ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations

M. Hutzenthaler and A. Jentzen and Th. Kruse and T. A. Nguyen

Research Report No. 2019-10
February 2019

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations

Martin Hutzenthaler[1], Arnulf Jentzen[2], Thomas Kruse[3] & Tuan Anh Nguyen[4]

[1] Faculty of Mathematics, University of Duisburg-Essen,
45117 Essen, Germany, e-mail: martin.hutzenthaler@uni-due.de

[2] Seminar for Applied Mathematics, Department of Mathematics,
ETH Zurich, Switzerland, e-mail: arnulf.jentzen@sam.math.ethz.ch

[3] Faculty of Mathematics, University of Duisburg-Essen,
45117 Essen, Germany, e-mail: thomas.kruse@uni-due.de

[4] Faculty of Mathematics, University of Duisburg-Essen,
45117 Essen, Germany, e-mail: tuan.nguyen@uni-due.de

Monday 4th February, 2019

Abstract

Deep neural networks and other deep learning methods have very successfully been applied to the numerical approximation of high-dimensional nonlinear parabolic partial differential equations (PDEs), which are widely used in finance, engineering, and natural sciences. In particular, simulations indicate that algorithms based on deep learning overcome the curse of dimensionality in the numerical approximation of solutions of semilinear PDEs. For certain linear PDEs this has also been proved mathematically. The key contribution of this article is to rigorously prove this for the first time for a class of nonlinear PDEs. More precisely, we prove in the case of semilinear heat equations with gradient-independent nonlinearities that the numbers of parameters of the employed deep neural networks grow at most polynomially in both the PDE dimension and the reciprocal of the prescribed approximation accuracy. Our proof relies on recently introduced multilevel Picard approximations of semilinear PDEs.

Contents

AMS 2010 subject classification: 65C99; 68T05
Key words and phrases: curse of dimensionality, high-dimensional PDEs, deep neural networks, information based complexity, tractability of multivariate problems, multilevel Picard approximations

1 Introduction

Deep neural networks (DNNs) have revolutionized a number of computational problems; see, e.g., the references in Grohs et al. [11]. In 2017 deep learning-based approximation algorithms for certain parabolic partial differential equations (PDEs) have been proposed in Han et al. [5, 12] and based on these works there is now a series of deep learning-based numerical approximation algorithms for a large class of different kinds of PDEs in the scientific literature; see, e.g., [1, 2, 3, 8, 9, 10, 11, 13, 17, 18, 19, 22, 23]. There is empirical evidence that deep learning-based methods work exceptionally well for approximating solutions of high-dimensional PDEs and that these do not suffer from the *curse of dimensionality*; see, e.g., the simulations in [5, 12, 2, 1]. There exist, however, only few theoretical results which prove that DNN approximations of solutions of PDEs do not suffer from the curse of dimensionality: The recent articles [11, 4, 16, 9] prove rigorously that DNN approximations overcome the curse of dimensionality in the numerical approximation of solutions of certain linear PDEs.

The key contribution of this article is to rigorously prove for the first time that DNN approximations overcome the curse of dimensionality in the numerical approximation of solutions of semilinear heat equations with gradient-independent nonlinearities.

Next we introduce our notation for DNNs. Throughout this article we use the so-called multilayer feedforward perceptron model which is a parametrized class of functions constructed by successive applications of affine mappings and coordinatewise nonlinearities (see Section 2 in Pinkus [21]), we use $\mathbb{R} \ni x \mapsto \max\{x, 0\} \in \mathbb{R}$ as activation function and $\mathbb{R}^d \ni x = (x_1, \ldots, x_d) \mapsto \mathbf{A}_d(x) = (\max\{x_1, 0\}, \ldots, \max\{x_d, 0\}) \in \mathbb{R}^d$, $d \in \mathbb{N}$, as DNN nonlinearities, and we follow the mathematical formulation of Peterson & Voigtlaender [20] and, especially, of Jentzen, Salimova & Welti [16]. The set of neural networks is denoted by

$$\mathcal{N} = \bigcup_{H \in \mathbb{N}} \bigcup_{(k_0, k_1, \ldots, k_H, k_{H+1}) \in \mathbb{N}^{H+2}} \prod_{n=1}^{H+1} \left(\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n} \right). \tag{1}$$

A neural network $\Phi = ((W_1, B_1), \ldots, (W_{H+1}, B_{H+1})) \in \prod_{n=1}^{H+1} \left(\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n} \right)$ with $H \in \mathbb{N}$ so-called hidden layers and vector $\mathcal{L}(\Phi) = (k_0, k_1, \ldots, k_{H+1}) \in \mathbb{N}^{H+2}$ of layer dimensions then defines a function $\mathcal{R}(\Phi) \in C(\mathbb{R}^{k_0}, \mathbb{R}^{k_{H+1}})$ which satisfies that for all $x_0 \in \mathbb{R}^{k_0}, \ldots, x_H \in \mathbb{R}^{k_H}$, $n \in \mathbb{N} \cap [1, H]$: $x_n = \mathbf{A}_{k_n}(W_n x_{n-1} + B_n)$ it holds that $(\mathcal{R}(\Phi))(x_0) = W_{H+1} x_H + B_{H+1}$. Moreover, the number of parameters of a neural network $\Phi \in \mathcal{N}$ is denoted by $\mathcal{P}(\Phi) \in \mathbb{N}$.

The main result of this article, Theorem 4.1 below, proves for semilinear heat equations with gradient-independent nonlinearities that the number of parameters of the approximating DNN grows at most polynomially in both the PDE dimension $d \in \mathbb{N}$ and the reciprocal of the prescribed accuracy $\varepsilon > 0$. Thereby we establish for the first time that there exist DNN approximations of solutions of such PDEs which indeed overcome the curse of dimensionality. To illustrate Theorem 4.1, we formulate the following special case of Theorem 4.1 using the above notation on DNNs and the notation from Subsection 1.1.

Theorem 1.1. *Let* $T, L \in (0, \infty)$, $B \in [0, \infty)$, $p, \tilde{p} \in \mathbb{N}$, $q \in \mathbb{N}$, $\alpha, \beta \in [0, \infty)$, $f \in C(\mathbb{R}, \mathbb{R})$, *for every* $d \in \mathbb{N}$ *let* $g_d \in C(\mathbb{R}^d, \mathbb{R})$, $u_d \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$, *for every* $d \in \mathbb{N}$ *let* $\nu_d \colon \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ *be a probability measure on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, *for every* $\varepsilon \in (0, 1]$, $d \in \mathbb{N}$ *let* $\Phi_\varepsilon^f, \Phi_\varepsilon^{g_d} \in \mathcal{N}$, *assume for all* $d \in \mathbb{N}$, $v, w \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$, $t \in (0, T)$ *that* $\mathcal{R}(\Phi_\varepsilon^f) \in C(\mathbb{R}, \mathbb{R})$, $\mathcal{R}(\Phi_\varepsilon^{g_d}) \in C(\mathbb{R}^d, \mathbb{R})$, $\left| (\mathcal{R}(\Phi_\varepsilon^f))(w) - (\mathcal{R}(\Phi_\varepsilon^f))(v) \right| \leq L |w - v|$, $\left| (\mathcal{R}(\Phi_\varepsilon^f))(0) \right| \leq B$, $|(\mathcal{R}(\Phi_\varepsilon^{g_d}))(x)| \leq B d^p (1 + \|x\|)^p$, $\left| f(v) - (\mathcal{R}(\Phi_\varepsilon^f))(v) \right| \leq \varepsilon B (1 + |v|^q)$, $|g_d(x) - (\mathcal{R}(\Phi_\varepsilon^{g_d}))(x)| \leq \varepsilon B d^p (1 + \|x\|)^{pq}$, $\max\left\{ \dim\left(\mathcal{L}\left(\Phi_\varepsilon^f \right) \right), \dim\left(\mathcal{L}\left(\Phi_\varepsilon^{g_d} \right) \right) \right\} \leq d^p \varepsilon^{-\beta} B$, $\max\left\{ \left\| \mathcal{L}(\Phi_\varepsilon^f) \right\|_\infty, \left\| \mathcal{L}(\Phi_\varepsilon^{g_d}) \right\|_\infty \right\} \leq d^p \varepsilon^{-\alpha} B$, $\left(\int_{\mathbb{R}^d} \|y\|^{2pq} \, \nu_d(dy) \right)^{1/(2pq)} \leq B d^{\tilde{p}}$, $\sup_{s \in [0,T]} \sup_{y \in \mathbb{R}^d} \left(\frac{|u_d(s,y)|}{1 + \|y\|^p} \right) < \infty$, $u_d(T, x) = g_d(x)$, *and*

$$(\tfrac{\partial}{\partial t} u_d)(t, x) + \tfrac{1}{2} (\Delta_x u_d)(t, x) + f(u_d(t, x)) = 0. \tag{2}$$

Then there exist $(\Psi_{d,\varepsilon})_{d\in\mathbb{N},\varepsilon\in(0,1]}\subseteq\mathcal{N}$, $\eta\in(0,\infty)$, $C\colon(0,1]\to(0,\infty)$ *such that for all* $d\in\mathbb{N}$, $\varepsilon\in(0,1]$, $\gamma\in(0,1]$ *it holds that*

$$\mathcal{P}(\Psi_{d,\varepsilon})\leq C(\gamma)d^{\eta}\varepsilon^{-(4+2\alpha+\beta+\gamma)}, \tag{3}$$

$\mathcal{R}(\Psi_{d,\varepsilon})\in C(\mathbb{R}^d,\mathbb{R})$, *and*

$$\left(\int_{\mathbb{R}^d}\left|u_d(0,x)-(\mathcal{R}(\Psi_{d,\varepsilon}))(x)\right|^2\nu_d(dx)\right)^{1/2}\leq\varepsilon. \tag{4}$$

Theorem 1.1 is an in immediate consequence of Theorem 4.1 below. In the manner of the proof of Theorem 3.14 in [11] and the proof of Theorem 6.3 in [16], the proof of Theorem 4.1 below uses probabilistic arguments on a suitable artificial probability space. Moreover, the proof of Theorem 4.1 relies on recently introduced (full history) multilevel Picard approximations which have been proved to overcome the curse of dimensionality in the numerical approximation of solutions of semilinear heat equations at single space-time points; see [6, 7, 15, 14]. A key step in our proof is that realizations of these random approximations can be represented by DNNs; see Lemma 3.10 below.

The remainder of this article is organized as follows. In Section 2 we provide auxiliary results on multilevel Picard approximations ensuring that these approximations are stable against perturbations in the nonlinearity f and the terminal condition g of the PDE (2). In Section 3 we show that multilevel Picard approximations can be represented by DNNs and we provide bounds for the number of parameters of the representing DNN. We use the results of Section 2 and Section 3 to prove the main result Theorem 4.1 in Section 4.

1.1 Notation

Let $\|\cdot\|\colon\bigcup_{d\in\mathbb{N}}\mathbb{R}^d\to[0,\infty)$ be the function that satisfies for all $d\in\mathbb{N}$, $x=(x_1,\ldots,x_d)\in\mathbb{R}^d$ that $\|x\|=\sqrt{\sum_{i=1}^d(x_i)^2}$. Let $\|\cdot\|_\infty\colon\bigcup_{d\in\mathbb{N}}\mathbb{R}^d\to[0,\infty)$ be the function that satisfies for all $d\in\mathbb{N}$, $x=(x_1,\ldots,x_d)\in\mathbb{R}^d$ that $\|x\|_\infty=\max_{i\in[1,d]\cap\mathbb{N}}|x_i|$.

2 A stability result for multilevel Picard approximations

Setting 2.1. *Let* $d\in\mathbb{N}$, $T,L,\delta,B\in(0,\infty)$, $p,q\in[1,\infty)$, $f_1,f_2\in C\big([0,T]\times\mathbb{R}^d\times\mathbb{R},\mathbb{R}\big)$, $g_1,g_2\in C(\mathbb{R}^d,\mathbb{R})$, *assume for all* $t\in[0,T]$, $x\in\mathbb{R}^d$, $w,v\in\mathbb{R}$, $i\in\{1,2\}$ *that*

$$|f_i(t,x,w)-f_i(t,x,v)|\leq L\,|w-v|, \tag{5}$$

$$\max\big\{|f_i(t,x,0)|,|g_i(x)|\big\}\leq B\,\big(1+\|x\|\big)^p, \tag{6}$$

and

$$\max\big\{|f_1(t,x,v)-f_2(t,x,v)|,|g_1(x)-g_2(x)|\big\}\leq\delta\,\Big(\big(1+\|x\|\big)^{pq}+|v|^q\Big), \tag{7}$$

let $F_1,F_2\colon C\big([0,T]\times\mathbb{R}^d,\mathbb{R}\big)\to C\big([0,T]\times\mathbb{R}^d,\mathbb{R}\big)$ *be the functions which satisfy for all* $v\in C\big([0,T]\times\mathbb{R}^d,\mathbb{R}\big)$, $t\in[0,T]$, $x\in\mathbb{R}^d$, $i\in\{1,2\}$ *that*

$$(F_i(v))(t,x)=f_i(t,x,v(t,x)), \tag{8}$$

let $(\Omega,\mathcal{F},\mathbb{P})$ *be a probability space, let* $\mathbf{W}\colon[0,T]\times\Omega\to\mathbb{R}^d$ *be a standard Brownian motion with continuous sample paths, let* $u_1,u_2\in C([0,T]\times\mathbb{R}^d,\mathbb{R})$, *assume for all* $i\in\{1,2\}$, $s\in[0,T]$, $x\in\mathbb{R}^d$ *that*

$$\mathbb{E}\bigg[\big|g_i\left(x+\mathbf{W}_{T-s}\right)\big|+\int_s^T\left|\left(F_i(u_i)\right)\left(t,x+\mathbf{W}_{t-s}\right)\right|\,dt\bigg]<\infty \tag{9}$$

and

$$u_i(s,x) = \mathbb{E}\bigg[g_i\left(x + \mathbf{W}_{T-s}\right) + \int_s^T \left(F_i(u_i)\right)\left(t, x + \mathbf{W}_{t-s}\right)\, dt \bigg], \tag{10}$$

let $\Theta = \bigcup_{n\in\mathbb{N}} \mathbb{Z}^n$, *let* $\mathfrak{u}^\theta\colon \Omega \to [0,1]$, $\theta \in \Theta$, *be independent random variables which are uniformly distributed on* $[0,1]$, *let* $\mathcal{U}^\theta\colon [0,T]\times\Omega \to [0,T]$, $\theta\in\Theta$, *satisfy for all* $t\in[0,T]$, $\theta\in\Theta$ *that* $\mathcal{U}^\theta_t = t + (T-t)\mathfrak{u}^\theta$, *let* $W^\theta\colon [0,T]\times\Omega\to\mathbb{R}^d$, $\theta\in\Theta$, *be independent standard Brownian motions, assume that* $(\mathfrak{u}^\theta)_{\theta\in\Theta}$, $(W^\theta)_{\theta\in\Theta}$, *and* $\mathbf{W}$ *are independent, and let* $U^\theta_{n,M}\colon [0,T]\times\mathbb{R}^d\times\Omega\to\mathbb{R}$, $n,M\in\mathbb{Z}$, $\theta\in\Theta$, *be functions such that for all* $n,M\in\mathbb{N}$, $\theta\in\Theta$, $t\in[0,T]$, $x\in\mathbb{R}^d$ *it holds that* $U^\theta_{-1,M}(t,x) = U^\theta_{0,M}(t,x) = 0$ *and*

$$\begin{aligned} U^\theta_{n,M}(t,x) &= \frac{1}{M^n}\sum_{i=1}^{M^n} g_2\left(x + W^{(\theta,0,-i)}_T - W^{(\theta,0,-i)}_t\right) \\ &+ \sum_{l=0}^{n-1}\frac{(T-t)}{M^{n-l}}\left[\sum_{i=1}^{M^{n-l}}\left(F_2\big(U^{(\theta,l,i)}_{l,M}\big) - \mathbb{1}_{\mathbb{N}}(l)F_2\big(U^{(\theta,-l,i)}_{l-1,M}\big)\right)\left(\mathcal{U}^{(\theta,l,i)}_t, x + W^{(\theta,l,i)}_{\mathcal{U}^{(\theta,l,i)}_t} - W^{(\theta,l,i)}_t\right)\right]. \end{aligned} \tag{11}$$

Lemma 2.2 (*q*-th moment of the exact solution)**.** *Assume Setting 2.1 and let* $x\in\mathbb{R}^d$, $i\in\{1,2\}$. *Then it holds that*

$$\sup_{t\in[0,T]}\left(\mathbb{E}\Big[|u_i(t,x+\mathbf{W}_t)|^q\Big]\right)^{1/q} \le e^{LT}(T+1)B\left[\sup_{t\in[0,T]}\left(\mathbb{E}\Big[\Big(1+\|x+\mathbf{W}_t\|\Big)^{pq}\Big]\right)^{1/q}\right]. \tag{12}$$

Proof of Lemma 2.2. Throughout this proof let $\mu_t\colon\mathcal{B}(\mathbb{R}^d)\to[0,1]$, $t\in[0,T]$ be the probability measures which satisfy for all $t\in[0,T]$, $B\in\mathcal{B}(\mathbb{R}^d)$ that

$$\mu_t(B) = \mathbb{P}(x+\mathbf{W}_t\in B). \tag{13}$$

The integral transformation theorem, (10), and the triangle inequality show for all $t\in[0,T]$ that

$$\begin{aligned} &\left(\mathbb{E}\Big[|u_i(t,x+\mathbf{W}_t)|^q\Big]\right)^{1/q} = \left(\int_{\mathbb{R}^d}|u_i(t,z)|^q\,\mu_t(dz)\right)^{1/q} \\ &= \left(\int_{\mathbb{R}^d}\left|\mathbb{E}\bigg[g_i(z+\mathbf{W}_{T-t}) + \int_t^T (F_i(u_i))(s,z+\mathbf{W}_{s-t})\,ds\bigg]\right|^q\,\mu_t(dz)\right)^{1/q} \\ &\le \left(\int_{\mathbb{R}^d}\big|\mathbb{E}\big[g_i(z+\mathbf{W}_{T-t})\big]\big|^q\,\mu_t(dz)\right)^{1/q} \\ &\quad + \int_t^T\left(\int_{\mathbb{R}^d}\big|\mathbb{E}\big[(F_i(u_i))(s,z+\mathbf{W}_{s-t})\big]\big|^q\,\mu_t(dz)\right)^{1/q} ds. \end{aligned} \tag{14}$$

Next, Jensen's inequality, Fubini's theorem, (13), the fact that $\mathbf{W}$ has independent and stationary increments, and (6) demonstrate that for all $t\in[0,T]$ it holds that

$$\begin{aligned} &\int_{\mathbb{R}^d}\big|\mathbb{E}\big[g_i(z+\mathbf{W}_{T-t})\big]\big|^q\,\mu_t(dz) \le \int_{\mathbb{R}^d}\mathbb{E}\Big[|g_i(z+\mathbf{W}_T-\mathbf{W}_t)|^q\Big]\,\mu_t(dz) \\ &= \mathbb{E}\Big[|g_i\left(x+\mathbf{W}_t+\mathbf{W}_T-\mathbf{W}_t\right)|^q\Big] = \mathbb{E}\Big[|g_i\left(x+\mathbf{W}_T\right)|^q\Big] \le \mathbb{E}\Big[B^q\left(1+\|x+\mathbf{W}_T\|\right)^{pq}\Big]. \end{aligned} \tag{15}$$

Furthermore, Jensen's inequality, Fubini's theorem, (13), the fact that $\mathbf{W}$ has independent and stationary increments, the triangle inequality, (5), and (6) demonstrate for all $t \in [0,T]$ that

$$\begin{aligned}
&\int_t^T \left(\int_{\mathbb{R}^d} \left| \mathbb{E}\big[(F_i(u_i))(s, z + \mathbf{W}_{s-t})\big] \right|^q \mu_t(dz) \right)^{1/q} ds \\
&\le \int_t^T \left(\int_{\mathbb{R}^d} \mathbb{E}\Big[\big|(F_i(u_i))(s, z + \mathbf{W}_s - \mathbf{W}_t)\big|^q \Big] \mu_t(dz) \right)^{1/q} ds \\
&= \int_t^T \left(\mathbb{E}\Big[\big| \left(F_i(u_i)\right)(s, x + \mathbf{W}_t + \mathbf{W}_s - \mathbf{W}_t) \big|^q \Big] \right)^{1/q} ds \\
&\le \int_t^T \left(\mathbb{E}\Big[|(F_i(0))(s, x + \mathbf{W}_s)|^q \Big] \right)^{1/q} ds + \int_t^T \left(\mathbb{E}\Big[|(F_i(u_i) - F_i(0))(s, x + \mathbf{W}_s)|^q \Big] \right)^{1/q} ds \\
&\le T \sup_{s \in [0,T]} \left(\mathbb{E}\Big[B^q \left(1 + \|x + \mathbf{W}_s\| \right)^{pq} \Big] \right)^{1/q} + \int_t^T \left(\mathbb{E}\Big[L^q \, |u_i(s, x + \mathbf{W}_s)|^q \Big] \right)^{1/q} ds.
\end{aligned} \tag{16}$$

Combining this with (14) and (15) implies that for all $t \in [0,T]$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}\Big[\big| u_i(t, x + \mathbf{W}_t) \big|^q \Big] \right)^{1/q} \\
&\le (T+1) B \sup_{s \in [0,T]} \left(\mathbb{E}\Big[\left(1 + \|x + \mathbf{W}_s\| \right)^{pq} \Big] \right)^{1/q} + L \int_t^T \left(\mathbb{E}\Big[|u_i(s, x + \mathbf{W}_s)|^q \Big] \right)^{1/q} ds.
\end{aligned} \tag{17}$$

Next, [14, Corollary 3.11] shows that

$$\sup_{s \in [0,T]} \sup_{y \in \mathbb{R}^d} \frac{|u_i(s,y)|}{(1 + \|y\|)^p} \le \sup_{s \in [0,T]} \sup_{y \in \mathbb{R}^d} \frac{|u_i(s,y)|}{1 + \|y\|^p} < \infty. \tag{18}$$

This, the triangle inequality, and the fact that $\mathbb{E}[\|\mathbf{W}_T\|^{pq}] < \infty$ show that

$$\begin{aligned}
&\int_0^T \left(\mathbb{E}\Big[|u_i(s, x + \mathbf{W}_s)|^q \Big] \right)^{1/q} ds \le \left[\sup_{s \in [0,T]} \sup_{y \in \mathbb{R}^d} \frac{|u(s,y)|}{(1 + \|y\|)^p} \right] \int_0^T \left(\mathbb{E}\Big[\left(1 + \|x + \mathbf{W}_s\|\right)^{pq} \Big] \right)^{1/q} ds \\
&\le \left[\sup_{s \in [0,T]} \sup_{y \in \mathbb{R}^d} \frac{|u(s,y)|}{(1 + \|y\|)^p} \right] T \left(1 + \|x\| + \left(\mathbb{E}\Big[\|\mathbf{W}_T\|^{pq} \Big] \right)^{\frac{1}{pq}} \right)^p < \infty.
\end{aligned} \tag{19}$$

This, Gronwall's integral inequality, and (17) establish for all $t \in [0,T]$ that

$$\left(\mathbb{E}\Big[|u_i(t, x + \mathbf{W}_t)|^q \Big] \right)^{1/q} \le e^{LT} (T+1) B \sup_{s \in [0,T]} \left(\mathbb{E}\Big[\left(1 + \|x + \mathbf{W}_s\| \right)^{pq} \Big] \right)^{1/q}. \tag{20}$$

The proof of Lemma 2.2 is thus completed. ☐

Lemma 2.3. *Assume Setting 2.1. Then it holds for all $t \in [0,T]$, $x \in \mathbb{R}^d$ that*

$$\begin{aligned}
&\mathbb{E}\Big[|u_1(t, x + \mathbf{W}_t) - u_2(t, x + \mathbf{W}_t)| \Big] \\
&\le \delta \left(e^{LT}(T+1) \right)^{q+1} (B^q + 1) \left(1 + \|x\| + \left(\mathbb{E}\Big[\|\mathbf{W}_T\|^{pq} \Big] \right)^{\frac{1}{pq}} \right)^{pq}.
\end{aligned} \tag{21}$$

Proof of Lemma 2.3. First, (10), the triangle inequality, and the fact that $\mathbf{W}$ has stationary increments show for all $s \in [0,T]$, $z \in \mathbb{R}^d$ that

$$\begin{aligned}
&|u_1(s,z) - u_2(s,z)| \\
&= \left|\mathbb{E}\left[(g_1 - g_2)(z + \mathbf{W}_{T-s}) + \int_s^T \big(F_1(u_1) - F_1(u_2) + F_1(u_2) - F_2(u_2)\big)(t, z + \mathbf{W}_{t-s})\, dt\right]\right| \\
&\leq \mathbb{E}\Big[\big|(g_1 - g_2)(z + \mathbf{W}_T - \mathbf{W}_s)\big|\Big] + \int_s^T \mathbb{E}\Big[\big|\big(F_1(u_1) - F_1(u_2)\big)(t, z + \mathbf{W}_t - \mathbf{W}_s)\big|\Big]\, dt \\
&\qquad + \int_s^T \mathbb{E}\Big[\big|\big(F_1(u_2) - F_2(u_2)\big)(t, z + \mathbf{W}_t - \mathbf{W}_s)\big|\Big]\, dt.
\end{aligned} \tag{22}$$

This, Fubini's theorem, the fact that $\mathbf{W}$ has independent increments, and the Lipschitz condition in (5) ensure that for all $s \in [0,T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\mathbb{E}\Big[\big|(u_1 - u_2)(s, x + \mathbf{W}_s)\big|\Big] = \mathbb{E}\Big[\big|u_1(s,z) - u_2(s,z)\big|\big|_{z=x+\mathbf{W}_s}\Big] \\
&\leq \mathbb{E}\left[\mathbb{E}\Big[\big|(g_1 - g_2)(z + \mathbf{W}_T - \mathbf{W}_s)\big|\Big]\Big|_{z=x+\mathbf{W}_s}\right] \\
&\qquad + \int_s^T \mathbb{E}\left[\mathbb{E}\Big[\big|\big(F_1(u_1) - F_1(u_2)\big)(t, z + \mathbf{W}_t - \mathbf{W}_s)\big|\Big]\Big|_{z=x+\mathbf{W}_s}\right] dt \\
&\qquad + \int_s^T \mathbb{E}\left[\mathbb{E}\Big[\big|\big(F_1(u_2) - F_2(u_2)\big)(t, z + \mathbf{W}_t - \mathbf{W}_s)\big|\Big]\Big|_{z=x+\mathbf{W}_s}\right] dt \\
&= \mathbb{E}\Big[\big|(g_1 - g_2)(x + \mathbf{W}_T)\big|\Big] + \int_s^T \mathbb{E}\Big[\big|\big(F_1(u_1) - F_1(u_2)\big)(t, x + \mathbf{W}_t)\big|\Big]\, dt \\
&\qquad + \int_s^T \mathbb{E}\Big[\big|\big(F_1(u_2) - F_2(u_2)\big)(t, x + \mathbf{W}_t)\big|\Big]\, dt \\
&\leq \mathbb{E}\Big[\big|(g_1 - g_2)(x + \mathbf{W}_T)\big|\Big] + \int_s^T \mathbb{E}\Big[L\,\big|(u_1 - u_2)(t, x + \mathbf{W}_t)\big|\Big]\, dt \\
&\qquad + T \sup_{t\in[0,T]} \mathbb{E}\Big[\big|\big(F_1(u_2) - F_2(u_2)\big)(t, x + \mathbf{W}_t)\big|\Big].
\end{aligned} \tag{23}$$

This, Gronwall's lemma, and Lemma 2.2 yield for all $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\sup_{t\in[0,T]} \mathbb{E}\Big[\big|(u_1 - u_2)(t, x + \mathbf{W}_t)\big|\Big] \\
&\leq e^{LT}(T+1) \sup_{t\in[0,T]} \max\Big\{\mathbb{E}\Big[\big|(g_1 - g_2)(x + \mathbf{W}_T)\big|\Big], \mathbb{E}\Big[\big|\big(F_1(u_2) - F_2(u_2)\big)(t, x + \mathbf{W}_t)\big|\Big]\Big\}.
\end{aligned} \tag{24}$$

Furthermore, (7), the triangle inequality, and Lemma 2.2 imply for all $x \in \mathbb{R}^d$ that

$$\begin{aligned}
&\sup_{t\in[0,T]} \max\Big\{\mathbb{E}\Big[\big|(g_1 - g_2)(x + \mathbf{W}_T)\big|\Big], \mathbb{E}\Big[\big|\big(F_1(u_2) - F_2(u_2)\big)(t, x + \mathbf{W}_t)\big|\Big]\Big\} \\
&\leq \delta \sup_{t\in[0,T]} \mathbb{E}\Big[\big(1 + \|x + \mathbf{W}_t\|\big)^{pq} + |u_2(x + \mathbf{W}_t)|^q\Big] \\
&\leq \delta \sup_{t\in[0,T]} \mathbb{E}\Big[\big(1 + \|x + \mathbf{W}_t\|\big)^{pq}\Big] + \delta \sup_{t\in[0,T]} \mathbb{E}\Big[|u_2(x + \mathbf{W}_t)|^q\Big]. \\
&\leq \delta \sup_{t\in[0,T]} \mathbb{E}\Big[\big(1 + \|x + \mathbf{W}_t\|\big)^{pq}\Big] + \delta(e^{LT}(T+1)B)^q \sup_{t\in[0,T]} \mathbb{E}\Big[\big(1 + \|x + \mathbf{W}_t\|\big)^{pq}\Big] \\
&\leq \delta \left(e^{LT}(T+1)\right)^q (B^q + 1) \sup_{t\in[0,T]} \mathbb{E}\Big[\big(1 + \|x + \mathbf{W}_t\|\big)^{pq}\Big].
\end{aligned} \tag{25}$$

This, (24), and the triangle inequality yield that

$$\begin{aligned}
&\sup_{t\in[0,T]} \mathbb{E}\Big[\big|(u_1-u_2)\,(t,x+\mathbf{W}_t)\big|\Big] \\
&\le \delta\left(e^{LT}(T+1)\right)^{q+1}(B^q+1)\sup_{t\in[0,T]}\mathbb{E}\left[\left(1+\|x+\mathbf{W}_t\|\right)^{pq}\right] \\
&\le \delta\left(e^{LT}(T+1)\right)^{q+1}(B^q+1)\left(1+\|x\|+\left(\mathbb{E}\Big[\|\mathbf{W}_T\|^{pq}\Big]\right)^{\frac{1}{pq}}\right)^{pq}.
\end{aligned} \tag{26}$$

This completes the proof of Lemma 2.3. □

Corollary 2.4. *Assume Setting 2.1, let $x\in\mathbb{R}^d$, $N,M\in\mathbb{N}$, and assume that $q\ge 2$. Then it holds that*

$$\begin{aligned}
&\left(\mathbb{E}\Big[\big|U^0_{N,M}(0,x)-u_1(0,x)\big|^2\Big]\right)^{1/2} \\
&\le \left(e^{LT}(T+1)\right)^{q+1}(B^q+1)\left(\delta+\frac{e^{M/2}(1+2LT)^N}{M^{N/2}}\right)\left(1+\|x\|+\left(\mathbb{E}\Big[\|\mathbf{W}_T\|^{pq}\Big]\right)^{\frac{1}{pq}}\right)^{pq}.
\end{aligned} \tag{27}$$

Proof of Corollary 2.4. First, Lemma 2.2 implies that $\int_0^T \left(\mathbb{E}\big[|u_i(t,x+\mathbf{W}_t)|^2\big]\right)^{1/2} dt<\infty$. This, [14, Theorem 3.5] (applied with $\xi=x$, $F=F_2$, $g=g_2$, and $u=u_2$ in the notation of [14, Theorem 3.5]), (6), and the triangle inequality ensure that

$$\begin{aligned}
&\left(\mathbb{E}\Big[\big|U^0_{N,M}(0,x)-u_2(0,x)\big|^2\Big]\right)^{1/2} \\
&\le e^{LT}\left[\left(\mathbb{E}\Big[\big|g_2(x+\mathbf{W}_T)\big|^2\Big]\right)^{1/2}+T\left(\frac{1}{T}\int_0^T\mathbb{E}\Big[\big|(F_2(0))(t,x+\mathbf{W}_t)\big|^2\Big]\,dt\right)^{1/2}\right]\frac{e^{M/2}(1+2LT)^N}{M^{N/2}} \\
&\le e^{LT}(T+1)\sup_{t\in[0,T]}\left(\mathbb{E}\left[B^2\left(1+\|x+\mathbf{W}_t\|\right)^{2p}\right]\right)^{1/2}\frac{e^{M/2}(1+2LT)^N}{M^{N/2}} \\
&\le e^{LT}(T+1)B\left(1+\|x\|+\left(\mathbb{E}\Big[\|\mathbf{W}_T\|^{2p}\Big]\right)^{\frac{1}{2p}}\right)^{p}\frac{e^{M/2}(1+2LT)^N}{M^{N/2}}.
\end{aligned} \tag{28}$$

Furthermore, Lemma 2.3 shows that

$$|u_2(0,x)-u_1(0,x)|\le\delta\left(e^{LT}(T+1)\right)^{q+1}(B^q+1)\left(1+\|x\|+\left(\mathbb{E}\Big[\|\mathbf{W}_T\|^{pq}\Big]\right)^{\frac{1}{pq}}\right)^{pq}. \tag{29}$$

This, the triangle inequality, (28), the fact that $B\le B^q+1$, the assumption that $q\ge 2$, and Jensen's inequality show that

$$\begin{aligned}
&\left(\mathbb{E}\Big[\big|U^0_{N,M}(0,x)-u_1(0,x)\big|^2\Big]\right)^{1/2} \\
&\le \left(\mathbb{E}\Big[\big|U^0_{N,M}(0,x)-u_2(0,x)\big|^2\Big]\right)^{1/2}+\big|u_2(0,x)-u_1(0,x)\big| \\
&\le \left(e^{LT}(T+1)\right)^{q+1}(B^q+1)\left(\delta+\frac{e^{M/2}(1+2LT)^N}{M^{N/2}}\right)\left(1+\|x\|+\left(\mathbb{E}\Big[\|\mathbf{W}_T\|^{pq}\Big]\right)^{\frac{1}{pq}}\right)^{pq}.
\end{aligned} \tag{30}$$

The proof of Corollary 2.4 is thus completed. □

3 Deep neural networks representing multilevel Picard approximations

The main result of this section, Lemma 3.10 below, shows that multilevel Picard aproximations can be well represented by DNNs. The central tools for the proof of Lemma 3.10 are Lemmas 3.8 and 3.9 which show that DNNs are stable under compositions and summations. We formulate Lemmas 3.8 and 3.9 in terms of the operators defined in (39) and (40) below, whose properties are studied in Lemmas 3.3, 3.4, and 3.5.

3.1 Results on deep neural networks

Setting 3.1 (Artificial neural networks)**.** *Let $\mathbf{A}_d\colon \mathbb{R}^d \to \mathbb{R}^d$, $d \in \mathbb{N}$, be the functions such that for all $d \in \mathbb{N}$, $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ it holds that*

$$\mathbf{A}_d(x) = (\max\{x_1, 0\}, \ldots, \max\{x_d, 0\}), \tag{31}$$

let $\mathcal{N}$ and $\mathcal{D}$ be the sets which satisfy that

$$\mathcal{N} = \bigcup_{H\in\mathbb{N}} \bigcup_{(k_0,k_1,\ldots,k_H,k_{H+1})\in\mathbb{N}^{H+2}} \prod_{n=1}^{H+1} \left(\mathbb{R}^{k_n\times k_{n-1}} \times \mathbb{R}^{k_n}\right) \quad \text{and} \quad \mathcal{D} = \bigcup_{H\in\mathbb{N}} \mathbb{N}^{H+2}, \tag{32}$$

let

$$\mathcal{L}\colon \mathcal{N} \to \mathcal{D} \quad \text{and} \quad \mathcal{R}\colon \mathcal{N} \to \bigcup_{k,l\in\mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l) \tag{33}$$

be the functions which satisfy that for all $H \in \mathbb{N}$, $k_0, k_1, \ldots, k_H, k_{H+1} \in \mathbb{N}$,

$$\Phi = ((W_1, B_1), \ldots, (W_{H+1}, B_{H+1})) \in \prod_{n=1}^{H+1} \left(\mathbb{R}^{k_n\times k_{n-1}} \times \mathbb{R}^{k_n}\right), \tag{34}$$

$x_0 \in \mathbb{R}^{k_0}, \ldots, x_H \in \mathbb{R}^{k_H}$ with $\forall n \in \mathbb{N} \cap [1, H]\colon x_n = \mathbf{A}_{k_n}(W_n x_{n-1} + B_n)$ it holds that

$$\begin{aligned} &\mathcal{L}(\Phi) = (k_0, k_1, \ldots, k_H, k_{H+1}), \quad \mathcal{R}(\Phi) \in C(\mathbb{R}^{k_0}, \mathbb{R}^{k_{H+1}}), \quad \text{and} \\ &(\mathcal{R}(\Phi))(x_0) = W_{H+1}x_H + B_{H+1}, \end{aligned} \tag{35}$$

let $\odot\colon \mathcal{D} \times \mathcal{D} \to \mathcal{D}$ be the binary operation with the property that for all $H_1, H_2 \in \mathbb{N}$,

$$\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_{H_1}, \alpha_{H_1+1}) \in \mathbb{N}^{H_1+2}, \quad \beta = (\beta_0, \beta_1, \ldots, \beta_{H_2}, \beta_{H_2+1}) \in \mathbb{N}^{H_2+2} \tag{36}$$

it holds that

$$\alpha \odot \beta = (\beta_0, \beta_1, \ldots, \beta_{H_2}, \beta_{H_2+1} + \alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{H_1+1}) \in \mathbb{N}^{H_1+H_2+3}, \tag{37}$$

let

$$\boxplus\colon \bigcup_{H,k,l\in\mathbb{N}} \left(\{k\} \times \mathbb{N}^H \times \{l\}\right)^2 \to \bigcup_{H,k,l\in\mathbb{N}} \left(\{k\} \times \mathbb{N}^H \times \{l\}\right) \tag{38}$$

be the function which satisfies that for all $H, k, l \in \mathbb{N}$,

$$\begin{aligned} \alpha &= (k, \alpha_1, \alpha_2, \ldots, \alpha_H, l) \in \{k\} \times \mathbb{N}^H \times \{l\}, \\ \beta &= (k, \beta_1, \beta_2, \ldots, \beta_H, l) \in \{k\} \times \mathbb{N}^H \times \{l\} \end{aligned} \tag{39}$$

it holds that

$$\alpha \boxplus \beta = (k, \alpha_1 + \beta_1, \ldots, \alpha_H + \beta_H, l) \in \{k\} \times \mathbb{N}^H \times \{l\}, \tag{40}$$

and let $\mathfrak{n}_n \in \mathcal{D}$, $n \in [3,\infty) \cap \mathbb{N}$, satisfy for all $n \in [3,\infty) \cap \mathbb{N}$ that

$$\mathfrak{n}_n = (1, \underbrace{2, \ldots, 2}_{(n-2)\text{-times}}, 1) \in \mathbb{N}^n. \tag{41}$$

Remark 3.2. *The set $\mathcal{N}$ can be viewed as the set of all artificial neural networks. For each network $\Phi \in \mathcal{N}$ the function $\mathcal{R}(\Phi)$ is the function represented by Φ and the vector $\mathcal{L}(\Phi)$ describes the layer dimensions of Φ.*

Lemma 3.3 ($\odot$ is associative)**.** *Assume Setting 3.1 and let $\alpha, \beta, \gamma \in \mathcal{D}$. Then it holds that $(\alpha \odot \beta) \odot \gamma = \alpha \odot (\beta \odot \gamma)$.*

Proof of Lemma 3.3. Throughout this proof let $H_1, H_2, H_3 \in \mathbb{N}$, let $(\alpha_i)_{i \in [0, H_1+1] \cap \mathbb{N}_0} \in \mathbb{N}^{H_1+2}$, $(\beta_i)_{i \in [0, H_2+1] \cap \mathbb{N}_0} \in \mathbb{N}^{H_2+2}$, $(\gamma_i)_{i \in [0, H_3+1] \cap \mathbb{N}_0} \in \mathbb{N}^{H_3+2}$ satisfy that

$$\begin{aligned} &\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_{H_1+1}), \quad \beta = (\beta_0, \beta_1, \ldots, \beta_{H_2+1}), \quad \text{and} \\ &\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_{H_3+1}). \end{aligned} \tag{42}$$

The definition of $\odot$ in (37) then shows that

$$\begin{aligned} (\alpha \odot \beta) \odot \gamma &= (\beta_0, \beta_1, \beta_2 \ldots, \beta_{H_2}, \beta_{H_2+1} + \alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{H_1+1}) \odot (\gamma_0, \gamma_1, \ldots, \gamma_{H_3+1}) \\ &= (\gamma_0, \ldots, \gamma_{H_3}, \gamma_{H_3+1} + \beta_0, \beta_1, \ldots, \beta_{H_2}, \beta_{H_2+1} + \alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{H_1+1}) \\ &= (\alpha_0, \alpha_1, \ldots, \alpha_{H_1+1}) \odot (\gamma_0, \gamma_1, \ldots, \gamma_{H_3}, \gamma_{H_3+1} + \beta_0, \beta_1, \beta_2, \ldots, \beta_{H_2+1}) \\ &= \alpha \odot (\beta \odot \gamma). \end{aligned} \tag{43}$$

The proof of Lemma 3.3 is thus completed. □

Lemma 3.4 ($\boxplus$ and associativity)**.** *Assume Setting 3.1, let $H, k, l \in \mathbb{N}$, and let $\alpha, \beta, \gamma \in \big(\{k\} \times \mathbb{N}^H \times \{l\}\big)$. Then*

i) it holds that $\alpha \boxplus \beta \in \big(\{k\} \times \mathbb{N}^H \times \{l\}\big)$,

ii) it holds that $\beta \boxplus \gamma \in \big(\{k\} \times \mathbb{N}^H \times \{l\}\big)$, and

iii) it holds that $(\alpha \boxplus \beta) \boxplus \gamma = \alpha \boxplus (\beta \boxplus \gamma)$.

Proof of Lemma 3.4. Throughout this proof let $\alpha_i, \beta_i, \gamma_i \in \mathbb{N}$, $i \in [1, H] \cap \mathbb{N}$, satisfy that $\alpha = (k, \alpha_1, \alpha_2, \ldots, \alpha_H, l)$, $\beta = (k, \beta_1, \beta_2, \ldots, \beta_H, l)$, and $\gamma = (k, \gamma_1, \gamma_2, \ldots, \gamma_H, l)$. The definition of $\boxplus$ (see (38)–(40)) then shows that

$$\begin{aligned} \alpha \boxplus \beta &= (k, \alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_H + \beta_H, l) \in \{k\} \times \mathbb{N}^H \times \{l\}, \\ \beta \boxplus \gamma &= (k, \beta_1 + \gamma_1, \beta_2 + \gamma_2, \ldots, \beta_H + \gamma_H, l) \in \{k\} \times \mathbb{N}^H \times \{l\}, \end{aligned} \tag{44}$$

and

$$\begin{aligned} (\alpha \boxplus \beta) \boxplus \gamma &= (k, (\alpha_1 + \beta_1) + \gamma_1, (\alpha_2 + \beta_2) + \gamma_2, \ldots, (\alpha_H + \beta_H) + \gamma_H, l) \\ &= (k, \alpha_1 + (\beta_1 + \gamma_1), \alpha_2 + (\beta_2 + \gamma_2), \ldots, \alpha_H + (\beta_H + \gamma_H), l) = \alpha \boxplus (\beta \boxplus \gamma). \end{aligned} \tag{45}$$

The proof of Lemma 3.4 is thus completed. □

Lemma 3.5 (Triangle inequality)**.** *Assume Setting 3.1, let $H, k, l \in \mathbb{N}$, and let $\alpha, \beta \in \{k\} \times \mathbb{N}^H \times \{l\}$. Then it holds that $\|\alpha \boxplus \beta\|_\infty \leq \|\alpha\|_\infty + \|\beta\|_\infty$.*

Proof of Lemma 3.5. Throughout this proof let $\alpha_i, \beta_i \in \mathbb{N}$, $i \in [1, H] \cap \mathbb{N}$ satisfy that $\alpha = (k, \alpha_1, \alpha_2, \ldots, \alpha_H, l)$ and $\beta = (k, \beta_1, \beta_2, \ldots, \beta_H, l)$. The definition of $\boxplus$ (see (38)–(40)) then shows that $\alpha \boxplus \beta = (k, \alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_H + \beta_H, l)$. This together with the triangle inequality implies that

$$\begin{aligned} \|\alpha \boxplus \beta\|_\infty &= \sup\{|k|, |\alpha_1 + \beta_1|, |\alpha_2 + \beta_2|, \ldots, |\alpha_H + \beta_H|, |l|\} \\ &\leq \sup\{|k|, |\alpha_1|, |\alpha_2|, \ldots, |\alpha_H|, |l|\} + \sup\{|k|, |\beta_1|, |\beta_2|, \ldots, |\beta_H|, |l|\} \\ &= \|\alpha\|_\infty + \|\beta\|_\infty. \end{aligned} \tag{46}$$

This completes the proof of Lemma 3.5. □

The following result, Lemma 3.6, is a variant of [16, Lemma 5.4].

Lemma 3.6 (Existence of DNNs with $H \in \mathbb{N}$ hidden layers for the identity in $\mathbb{R}$)**.** *Assume Setting 3.1 and let $H \in \mathbb{N}$. Then it holds that* $\mathrm{Id}_{\mathbb{R}} \in \mathcal{R}(\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \mathfrak{n}_{H+2}\})$.

Proof of Lemma 3.6. Throughout this proof let $W_1 \in \mathbb{R}^{2\times 1}$, $W_i \in \mathbb{R}^{2\times 2}$, $i \in [2, H] \cap \mathbb{N}$, $W_{H+1} \in \mathbb{R}^{1\times 2}$, $B_i \in \mathbb{R}^2$, $i \in [1, H] \cap \mathbb{N}$, $B_{H+1} \in \mathbb{R}^1$ satisfy that

$$\begin{aligned} &W_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \forall i \in [2, H] \cap \mathbb{N} \colon W_i = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad W_{H+1} = \begin{pmatrix} 1 & -1 \end{pmatrix}, \\ &\forall i \in [1, H] \cap \mathbb{N} \colon B_i = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad B_{H+1} = 0, \end{aligned} \tag{47}$$

let $\phi \in \mathcal{N}$ satisfy that $\phi = ((W_1, B_1), (W_2, B_2), \ldots, (W_H, B_H), (W_{H+1}, B_{H+1}))$, for every $a \in \mathbb{R}$ let $a^+ \in [0, \infty)$ be the non-negative part of a, i.e., $a^+ = \max\{a, 0\}$, and let $x_0 \in \mathbb{R}$, $x_1, x_2, \ldots, x_H \in \mathbb{R}^2$ satisfy for all $n \in \mathbb{N} \cap [1, H]$ that

$$x_n = \mathbf{A}_2(W_n x_{n-1} + B_n). \tag{48}$$

Note that (47) and the definition of $\mathcal{L}$ (see (33)–(35)) imply that $\mathcal{L}(\phi) = \mathfrak{n}_{H+2}$. Furthermore, (47), (48), and an induction argument show that

$$\begin{aligned} x_1 &= \mathbf{A}_2(W_1 x_0 + B_1) = \mathbf{A}_2\left(\begin{pmatrix} x_0 \\ -x_0 \end{pmatrix}\right) = \begin{pmatrix} x_0^+ \\ (-x_0)^+ \end{pmatrix}, \\ x_2 &= \mathbf{A}_2(W_2 x_1 + B_2) = \mathbf{A}_2(x_1) = \mathbf{A}_2\left(\begin{pmatrix} x_0^+ \\ (-x_0)^+ \end{pmatrix}\right) = \begin{pmatrix} x_0^+ \\ (-x_0)^+ \end{pmatrix}, \\ &\vdots \\ x_H &= \mathbf{A}_2(W_H x_{H-1} + B_H) = \mathbf{A}_2(x_{H-1}) = \mathbf{A}_2\left(\begin{pmatrix} x_0^+ \\ (-x_0)^+ \end{pmatrix}\right) = \begin{pmatrix} x_0^+ \\ (-x_0)^+ \end{pmatrix}. \end{aligned} \tag{49}$$

The definition of $\mathcal{R}$ (see (33)–(35)) hence ensures that

$$(\mathcal{R}(\phi))(x_0) = W_{H+1} x_H + B_{H+1} = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} x_0^+ \\ (-x_0)^+ \end{pmatrix} = x_0^+ - (-x_0)^+ = x_0. \tag{50}$$

The fact that x_0 was arbitrary therefore proves that $\mathcal{R}(\phi) = \mathrm{Id}_{\mathbb{R}}$. This and the fact that $\mathcal{L}(\phi) = \mathfrak{n}_{H+2}$ demonstrate that $\mathrm{Id}_{\mathbb{R}} \in \mathcal{R}(\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \mathfrak{n}_{H+2}\})$. The proof of Lemma 3.6 is thus completed. ☐

Lemma 3.7 (DNNs for affine transformations)**.** *Assume Setting 3.1 and let $d, m \in \mathbb{N}$, $\lambda \in \mathbb{R}$, $b \in \mathbb{R}^d$, $a \in \mathbb{R}^m$, $\Psi \in \mathcal{N}$ satisfy that $\mathcal{R}(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^m)$. Then it holds that*

$$\lambda\Big(\big(\mathcal{R}(\Psi)\big)(\cdot + b) + a\Big) \in \mathcal{R}\Big(\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \mathcal{L}(\Psi)\}\Big). \tag{51}$$

Proof of Lemma 3.7. Throughout this proof let $H, k_0, k_1, \ldots, k_{H+1} \in \mathbb{N}$ satisfy that

$$H + 2 = \dim(\mathcal{L}(\Psi)) \quad \text{and} \quad (k_0, k_1, \ldots, k_H, k_{H+1}) = \mathcal{L}(\Psi), \tag{52}$$

let $((W_1, B_1), (W_2, B_2), \ldots, (W_H, B_H), (W_{H+1}, B_{H+1})) \in \prod_{n=1}^{H+1} \left(\mathbb{R}^{k_n \times k_{n-1}} \times \mathbb{R}^{k_n}\right)$ satisfy that

$$\Big((W_1, B_1), (W_2, B_2), \ldots, (W_H, B_H), (W_{H+1}, B_{H+1})\Big) = \Psi, \tag{53}$$

let $\phi \in \mathcal{N}$ satisfy that

$$\phi = \Big((W_1, B_1 + W_1 b), (W_2, B_2), \ldots, (W_H, B_H), (\lambda W_{H+1}, \lambda B_{H+1} + \lambda a)\Big), \tag{54}$$

and let $x_0, y_0 \in \mathbb{R}^{k_0}, x_1, y_1 \in \mathbb{R}^{k_1}, \ldots, x_H, y_H \in \mathbb{R}^{k_H}$ satisfy for all $n \in \mathbb{N} \cap [1, H]$ that

$$x_n = \mathbf{A}_{k_n}(W_n x_{n-1} + B_n),\ y_n = \mathbf{A}_{k_n}(W_n y_{n-1} + B_n + \mathbb{1}_{\{1\}}(n) W_1 b) \quad \text{and} \quad x_0 = y_0 + b. \tag{55}$$

Then it holds that

$$y_1 = \mathbf{A}_{k_1}(W_1 y_0 + B_1 + W_1 b) = \mathbf{A}_{k_1}(W_1(y_0 + b) + B_1) = \mathbf{A}_{k_1}(W_1 x_0 + B_1) = x_1. \tag{56}$$

This and an induction argument prove for all $i \in [2, H] \cap \mathbb{N}$ that

$$y_i = \mathbf{A}_{k_i}(W_i y_{i-1} + B_i) = \mathbf{A}_{k_i}(W_i x_{i-1} + B_i) = x_i. \tag{57}$$

The definition of $\mathcal{R}$ (see (33)–(35)) hence shows that

$$\begin{aligned} (\mathcal{R}(\phi))(y_0) &= \lambda W_{H+1} y_H + \lambda B_{H+1} + \lambda a = \lambda W_{H+1} x_H + \lambda B_{H+1} + \lambda a \\ &= \lambda(W_{H+1} x_H + B_{H+1} + a) = \lambda((\mathcal{R}(\Psi))(x_0) + a) = \lambda(\mathcal{R}(\Psi))(y_0 + b) + a. \end{aligned} \tag{58}$$

This and the fact that y_0 was arbitrary prove that $\mathcal{R}(\phi) = \lambda((\mathcal{R}(\Psi))(\cdot + b) + a)$. This and the fact that $\mathcal{L}(\phi) = \mathcal{L}(\Psi)$ imply that $\lambda\left((\mathcal{R}(\Psi))(\cdot + b) + a\right) \in \mathcal{R}(\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \mathcal{L}(\Psi)\})$. The proof of Lemma 3.7 is thus completed. □

Lemma 3.8 (Composition)**.** *Assume Setting 3.1 and let $d_1, d_2, d_3 \in \mathbb{N}$, $f \in C(\mathbb{R}^{d_2}, \mathbb{R}^{d_3})$, $u \in C(\mathbb{R}^{d_1}, \mathbb{R}^{d_2})$, $\alpha, \beta \in \mathcal{D}$ satisfy that $f \in \mathcal{R}(\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \alpha\})$ and $u \in \mathcal{R}(\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \beta\})$. Then it holds that $(f \circ u) \in \mathcal{R}(\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \alpha \odot \beta\})$.*

Proof of Lemma 3.8. Throughout this proof let $H_1, H_2, \alpha_0, \ldots, \alpha_{H_1+1}, \beta_0, \ldots, \beta_{H_2+1} \in \mathbb{N}, \Phi_f, \Phi_u \in \mathcal{N}$ satisfy that

$$\begin{aligned} &(\alpha_0, \alpha_1, \ldots, \alpha_{H_1+1}) = \alpha, \quad (\beta_0, \beta_1, \ldots, \beta_{H_2+1}) = \beta, \quad \mathcal{R}(\Phi_f) = f, \\ &\mathcal{L}(\Phi_f) = \alpha, \quad \mathcal{R}(\Phi_u) = u, \quad \text{and} \quad \mathcal{L}(\Phi_u) = \beta. \end{aligned} \tag{59}$$

Lemma 5.4 in [16] shows that there exists $\mathbb{I} \in \mathcal{N}$ such that $\mathcal{L}(\mathbb{I}) = d_2 \mathfrak{n}_3 = (d_2, 2d_2, d_2)$ and $\mathcal{R}(\mathbb{I}) = \mathrm{Id}_{\mathbb{R}^{d_2}}$. Note that $2d_2 = \beta_{H_2+1} + \alpha_0$. This and [16, Proposition 5.2] (with $\phi_1 = \Phi_f$, $\phi_2 = \Phi_u$, and $\mathbb{I} = \mathbb{I}$ in the notation of [16, Proposition 5.2]) show that there exists $\Phi_{f \circ u} \in \mathcal{N}$ such that $\mathcal{R}(\Phi_{f \circ u}) = f \circ u$ and $\mathcal{L}(\Phi_{f \circ u}) = \mathcal{L}(\Phi_f) \odot \mathcal{L}(\Phi_u) = \alpha \odot \beta$. Hence, it holds that $f \circ u \in \mathcal{R}(\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \alpha \odot \beta\})$. The proof of Lemma 3.8 is thus completed. □

The following result, Lemma 3.9, essentially generalizes [16, Lemma 5.1] to the case where the DNNs have different hidden layer dimensions.

Lemma 3.9 (Sum of DNNs of the same length)**.** *Assume Setting 3.1 and let $M, H, p, q \in \mathbb{N}$, $h_1, h_2, \ldots, h_M \in \mathbb{R}$, $k_i \in \mathcal{D}$, $f_i \in C(\mathbb{R}^p, \mathbb{R}^q)$, $i \in [1, M] \cap \mathbb{N}$, satisfy for all $i \in [1, M] \cap \mathbb{N}$ that*

$$\dim(k_i) = H + 2 \quad \textit{and} \quad f_i \in \mathcal{R}\left(\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = k_i\}\right). \tag{60}$$

Then it holds that

$$\sum_{i=1}^{M} h_i f_i \in \mathcal{R}\left(\left\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \mathop{\boxplus}_{i=1}^{M} k_i\right\}\right). \tag{61}$$

Proof of Lemma 3.9. Throughout this proof let $\phi_i \in \mathcal{N}$, $i \in [1, M] \cap \mathbb{N}$, and $k_{i,n} \in \mathbb{N}$, $i \in [1, M] \cap \mathbb{N}$, $n \in [0, H+1] \cap \mathbb{N}_0$, satisfy for all $i \in [1, M] \cap \mathbb{N}$ that

$$\mathcal{L}(\phi_i) = k_i = (k_{i,0}, k_{i,1}, k_{i,2}, \ldots, k_{i,H}, k_{i,H+1}) \quad \text{and} \quad \mathcal{R}(\phi_i) = f_i, \tag{62}$$

for every $i \in [1, M] \cap \mathbb{N}$ let $((W_{i,1}, B_{i,1}), \ldots, (W_{i,H+1}, B_{i,H+1})) \in \prod_{n=1}^{H+1} \left(\mathbb{R}^{k_{i,n} \times k_{i,n-1}} \times \mathbb{R}^{k_{i,n}}\right)$ satisfy that

$$\phi_i = ((W_{i,1}, B_{i,1}), \ldots, (W_{i,H+1}, B_{i,H+1})), \tag{63}$$

let $k_n^{\boxplus} \in \mathbb{N}$, $n \in [1, H] \cap \mathbb{N}$, $k^{\boxplus} \in \mathbb{N}^{H+2}$ satisfy for all $n \in [1, H] \cap \mathbb{N}$ that

$$k_n^{\boxplus} = \sum_{i=1}^{M} k_{i,n} \quad \text{and} \quad k^{\boxplus} = (p, k_1^{\boxplus}, k_2^{\boxplus}, \ldots, k_H^{\boxplus}, q), \tag{64}$$

let $W_1 \in \mathbb{R}^{k_1^{\boxplus} \times p}$, $B_1 \in \mathbb{R}^{k_1^{\boxplus}}$ satisfy that

$$W_1 = \begin{pmatrix} W_{1,1} \\ W_{2,1} \\ \vdots \\ W_{M,1} \end{pmatrix} \quad \text{and} \quad B_1 = \begin{pmatrix} B_{1,1} \\ B_{2,1} \\ \vdots \\ B_{M,1} \end{pmatrix}, \tag{65}$$

let $W_n \in \mathbb{R}^{k_n^{\boxplus} \times k_{n-1}^{\boxplus}}$, $B_n \in \mathbb{R}^{k_n^{\boxplus}}$, $n \in [2, H] \cap \mathbb{N}$, satisfy for all $n \in [2, H] \cap \mathbb{N}$ that

$$W_n = \begin{pmatrix} W_{1,n} & 0 & 0 & 0 \\ 0 & W_{2,n} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & W_{M,n} \end{pmatrix} \quad \text{and} \quad B_n = \begin{pmatrix} B_{1,n} \\ B_{2,n} \\ \vdots \\ B_{M,n} \end{pmatrix}, \tag{66}$$

let $W_{H+1} \in \mathbb{R}^{q \times k_H^{\boxplus}}$, $B_{H+1} \in \mathbb{R}^q$ satisfy that

$$W_{H+1} = \begin{pmatrix} h_1 W_{1,H+1} & \ldots & h_M W_{M,H+1} \end{pmatrix} \quad \text{and} \quad B_{H+1} = \sum_{i=1}^{M} h_i B_{i,H+1}, \tag{67}$$

let $x_0 \in \mathbb{R}^p$, $x_1 \in \mathbb{R}^{k_1^{\boxplus}}, x_2 \in \mathbb{R}^{k_2^{\boxplus}} \ldots, x_H \in \mathbb{R}^{k_H^{\boxplus}}$, let $x_{1,0}, x_{2,0}, \ldots, x_{M,0} \in \mathbb{R}^p$, $x_{i,n} \in \mathbb{R}^{k_{i,n}}$, $i \in [1, M] \cap \mathbb{N}$, $n \in [1, H] \cap \mathbb{N}$, satisfy for all $i \in [1, M] \cap \mathbb{N}$, $n \in [1, H] \cap \mathbb{N}$ that

$$\begin{aligned} &x_0 = x_{1,0} = x_{2,0} = \ldots = x_{M,0}, \\ &x_{i,n} = \mathbf{A}_{k_{i,n}}(W_{i,n} x_{i,n-1} + B_{i,n}), \\ &x_n = \mathbf{A}_{k_n^{\boxplus}}(W_n x_{n-1} + B_n), \end{aligned} \tag{68}$$

and let $\psi \in \mathcal{N}$ satisfy that

$$\psi = ((W_1, B_1), (W_2, B_2), \ldots, (W_H, B_H), (W_{H+1}, B_{H+1})). \tag{69}$$

First, the definitions of $\mathcal{L}$ and $\mathcal{R}$ (see (33)–(35)), (62), and the fact that $\forall i \in [1, M] \cap \mathbb{N}\colon f_i \in C(\mathbb{R}^p, \mathbb{R}^q)$ show for all $i \in [1, M] \cap \mathbb{N}$ that $k_i = (p, k_{i,1}, k_{i,2}, \ldots, k_{i,H}, q)$. The definition of $\mathcal{L}$ (see (33)–(35)), the definition of $\boxplus$ (see (38)–(40)), and (64) then show that

$$\mathcal{L}(\psi) = (p, k_1^{\boxplus}, \ldots, k_H^{\boxplus}, q) = \mathop{\boxplus}_{i=1}^{M} k_i. \tag{70}$$

Next, we prove by induction on $n \in [1, H] \cap \mathbb{N}$ that $x_n = (x_{1,n}, x_{2,n}, \ldots, x_{M,n})$. First, (65) shows that

$$W_1 x_0 + B_1 = \begin{pmatrix} W_{1,1} \\ W_{2,1} \\ \vdots \\ W_{M,1} \end{pmatrix} x_0 + \begin{pmatrix} B_{1,1} \\ B_{2,1} \\ \vdots \\ B_{M,1} \end{pmatrix} = \begin{pmatrix} W_{1,1} x_0 + B_{1,1} \\ W_{2,1} x_0 + B_{2,1} \\ \vdots \\ W_{M,1} x_0 + B_{M,1} \end{pmatrix}. \tag{71}$$

This implies that

$$x_1 = \mathbf{A}_{k_1^{\boxplus}}(W_1 x_0 + B_1) = \begin{pmatrix} x_{1,1} \\ x_{2,1} \\ \vdots \\ x_{M,1} \end{pmatrix}. \tag{72}$$

This proves the base case. Next, for the induction step let $n \in [2, H] \cap \mathbb{N}$ and assume that $x_{n-1} = (x_{1,n-1}, x_{2,n-1}, \ldots, x_{M,n-1})$. Then (66) and the induction hypothesis ensure that

$$\begin{aligned} & W_n x_{n-1} + B_n \\ & = W_n \begin{pmatrix} x_{1,n-1} \\ x_{2,n-1} \\ \vdots \\ x_{M,n-1} \end{pmatrix} + B_n = \begin{pmatrix} W_{1,n} & 0 & 0 & 0 \\ 0 & W_{2,n} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & W_{M,n} \end{pmatrix} \begin{pmatrix} x_{1,n-1} \\ x_{2,n-1} \\ \vdots \\ x_{M,n-1} \end{pmatrix} + \begin{pmatrix} B_{1,n} \\ B_{2,n} \\ \vdots \\ B_{M,n} \end{pmatrix} \\ & = \begin{pmatrix} W_{1,n} x_{1,n-1} + B_{1,n} \\ W_{2,n} x_{2,n-1} + B_{2,n} \\ \vdots \\ W_{M,n} x_{M,n-1} + B_{M,n} \end{pmatrix}. \end{aligned} \tag{73}$$

This yields that

$$x_n = \mathbf{A}_{k_n^{\boxplus}}(W_n x_{n-1} + B_n) = \begin{pmatrix} x_{1,n} \\ x_{2,n} \\ \vdots \\ x_{M,n} \end{pmatrix}. \tag{74}$$

This proves the induction step. Induction now proves for all $n \in [1, H] \cap \mathbb{N}$ that $x_n = (x_{1,n}, x_{2,n}, \ldots, x_{M,n})$. This, the definition of $\mathcal{R}$ (see (33)–(35)), and (67) imply that

$$\begin{aligned} & (\mathcal{R}(\psi))(x_0) = W_{H+1} x_H + B_{H+1} \\ & = W_{H+1} \begin{pmatrix} x_{1,H} \\ x_{2,H} \\ \vdots \\ x_{M,H} \end{pmatrix} + B_{H+1} = \begin{pmatrix} h_1 W_{1,H+1} & \ldots & h_M W_{M,H+1} \end{pmatrix} \begin{pmatrix} x_{1,H} \\ x_{2,H} \\ \vdots \\ x_{M,H} \end{pmatrix} + \left[\sum_{i=1}^{M} h_i B_{i,H+1} \right] \\ & = \left[\sum_{i=1}^{M} h_i W_{i,H+1} x_{i,H} \right] + \left[\sum_{i=1}^{M} h_i B_{i,H+1} \right] = \sum_{i=1}^{M} h_i \left(W_{i,H+1} x_{i,H} + B_{i,H+1} \right) \\ & = \sum_{i=1}^{M} h_i (\mathcal{R}(\phi_i))(x_0). \end{aligned} \tag{75}$$

This, the fact that $x_0 \in \mathbb{R}^p$ was arbitrary, and (62) yield that

$$\mathcal{R}(\psi) = \sum_{i=1}^{M} h_i \mathcal{R}(\phi_i) = \sum_{i=1}^{M} h_i f_i. \tag{76}$$

This and (70) show that

$$\sum_{i=1}^{M} h_i f_i \in \mathcal{R}\left(\left\{ \Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \mathop{\boxplus}_{i=1}^{M} k_i \right\} \right). \tag{77}$$

The proof of Lemma 3.9 is thus completed. ☐

3.2 Multilevel Picard approximations

Lemma 3.10. *Assume Setting 3.1, let* $d, M \in \mathbb{N}$, $T, c \in (0,\infty)$, $f \in C(\mathbb{R},\mathbb{R})$, $g \in C(\mathbb{R}^d,\mathbb{R})$, $\Phi_f, \Phi_g \in \mathcal{N}$ *satisfy that* $\mathcal{R}(\Phi_f) = f$, $\mathcal{R}(\Phi_g) = g$, *and*

$$c \geq \max\left\{2, \|\mathcal{L}(\Phi_f)\|_\infty, \|\mathcal{L}(\Phi_g)\|_\infty\right\}, \tag{78}$$

let $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, let* $\Theta = \bigcup_{n\in\mathbb{N}} \mathbb{Z}^n$, *let* $\mathfrak{u}^\theta \colon \Omega \to [0,1]$, $\theta \in \Theta$, *be independent random variables which are uniformly distributed on* $[0,1]$, *let* $\mathcal{U}^\theta \colon [0,T] \times \Omega \to [0,T]$, $\theta \in \Theta$, *satisfy for all* $t \in [0,T]$, $\theta \in \Theta$ *that* $\mathcal{U}^\theta_t = t + (T-t)\mathfrak{u}^\theta$, *let* $W^\theta \colon [0,T] \times \Omega \to \mathbb{R}^d$, $\theta \in \Theta$, *be independent standard Brownian motions with continuous sample paths, assume that* $(\mathfrak{u}^\theta)_{\theta\in\Theta}$ *and* $(W^\theta)_{\theta\in\Theta}$ *are independent, let* $U^\theta_{n,M} \colon [0,T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$, $n, M \in \mathbb{Z}$, $\theta \in \Theta$, *be functions such that for all* $n \in \mathbb{N}$, $\theta \in \Theta$, $t \in [0,T]$, $x \in \mathbb{R}^d$ *it holds that* $U^\theta_{-1,M}(t,x) = U^\theta_{0,M}(t,x) = 0$ *and*

$$\begin{aligned} U^\theta_{n,M}(t,x) &= \frac{1}{M^n} \sum_{i=1}^{M^n} g\big(x + W^{(\theta,0,-i)}_T - W^{(\theta,0,-i)}_t\big) \\ &+ \sum_{l=0}^{n-1} \frac{(T-t)}{M^{n-l}} \left[\sum_{i=1}^{M^{n-l}} \big(f \circ U^{(\theta,l,i)}_{l,M} - \mathbb{1}_{\mathbb{N}}(l) f \circ U^{(\theta,-l,i)}_{l-1,M}\big)\Big(\mathcal{U}^{(\theta,l,i)}_t, x + W^{(\theta,l,i)}_{\mathcal{U}^{(\theta,l,i)}_t} - W^{(\theta,l,i)}_t\Big)\right], \end{aligned} \tag{79}$$

and let $\omega \in \Omega$. *Then for all* $n \in \mathbb{N}_0$ *there exists a family* $(\Phi^\theta_{n,t})_{\theta\in\Theta, t\in[0,T]} \subseteq \mathcal{N}$ *such that*

i) it holds for all $t_1, t_2 \in [0,T]$, $\theta_1, \theta_2 \in \Theta$ *that*

$$\mathcal{L}\left(\Phi^{\theta_1}_{n,t_1}\right) = \mathcal{L}\left(\Phi^{\theta_2}_{n,t_2}\right), \tag{80}$$

ii) it holds for all $t \in [0,T]$, $\theta \in \Theta$ *that*

$$\begin{aligned} &\dim\left(\mathcal{L}\left(\Phi^\theta_{n,t}\right)\right) = n\Big(\dim\left(\mathcal{L}\left(\Phi_f\right)\right) - 1\Big) + \dim\left(\mathcal{L}\left(\Phi_g\right)\right) \quad \textit{and} \\ &\left\|\mathcal{L}(\Phi^\theta_{n,t})\right\|_\infty \leq c(3M)^n, \end{aligned} \tag{81}$$

and

iii) it holds for all $\theta \in \Theta$, $t \in [0,T]$ *that*

$$U^\theta_{n,M}(t,\cdot,\omega) = \mathcal{R}(\Phi^\theta_{n,t}). \tag{82}$$

Proof of Lemma 3.10. We prove Lemma 3.10 by induction on $n \in \mathbb{N}_0$. For the base case $n = 0$ note that the fact that $\forall t \in [0,T], \theta \in \Theta \colon U^\theta_{0,M}(t,\cdot) = 0$, the fact that the function 0 can be represented by a network with depth $\dim\left(\mathcal{L}\left(\Phi_g\right)\right)$, and (78) imply that there exists $(\Phi^\theta_{0,t})_{\theta\in\Theta,t\in[0,T]} \subseteq \mathcal{N}$ such that it holds for all $t_1, t_2 \in [0,T]$, $\theta_1, \theta_2 \in \Theta$ that $\mathcal{L}\left(\Phi^{\theta_1}_{0,t_1}\right) = \mathcal{L}\left(\Phi^{\theta_2}_{0,t_2}\right)$ and such that it holds for all $\theta \in \Theta$, $t \in [0,T]$ that $\dim\left(\mathcal{L}(\Phi^\theta_{0,t})\right) = \dim\left(\mathcal{L}\left(\Phi_g\right)\right)$, $\left\|\mathcal{L}(\Phi^\theta_{0,t})\right\|_\infty \leq \|\mathcal{L}(\Phi_g)\|_\infty \leq c$, and $U^\theta_{0,M}(t,\cdot,\omega) = \mathcal{R}(\Phi^\theta_{0,t})$. This proves the base case $n = 0$.

For the induction step from $n \in \mathbb{N}_0$ to $n+1 \in \mathbb{N}$ let $n \in \mathbb{N}_0$ and assume that Item (i)–Item (iii) hold true for all $k \in [0,n] \cap \mathbb{N}_0$. The assumption that $g = \mathcal{R}(\Phi_g)$ and Lemma 3.7 (applied with $d = d$, $m = 1$, $\lambda = 1$, $a = 0$, $b = W^\theta_T(\omega) - W^\theta_t(\omega)$, and $\Psi = \Phi_g$ for $\theta \in \Theta$, $t \in [0,T]$ in the notation of Lemma 3.7) show for all $\theta \in \Theta$, $t \in [0,T]$ that

$$\begin{aligned} g\big(\cdot + W^\theta_T(\omega) - W^\theta_t(\omega)\big) &= (\mathcal{R}(\Phi_g))\big(\cdot + W^\theta_T(\omega) - W^\theta_t(\omega)\big) \\ &\in \mathcal{R}\left(\left\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \mathcal{L}(\Phi_g)\right\}\right). \end{aligned} \tag{83}$$

Furthermore, Lemma 3.6 (applied with $H = (n+1)\left(\dim\left(\mathcal{L}(\Phi_f)\right) - 1\right) - 1$ in the notation of Lemma 3.6) ensures that

$$\mathrm{Id}_{\mathbb{R}} \in \mathcal{R}\left(\left\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \mathfrak{n}_{(n+1)\left(\dim\left(\mathcal{L}(\Phi_f)\right)-1\right)+1}\right\}\right). \tag{84}$$

This, (83), and Lemma 3.8 (applied with $d_1 = d$, $d_2 = 1$, $d_3 = 1$, $f = \mathrm{Id}_{\mathbb{R}}$, $u = g\big(\cdot + W^\theta_T(\omega) - W^\theta_t(\omega)\big)$, $\alpha = \mathfrak{n}_{(n+1)\left(\dim\left(\mathcal{L}(\Phi_f)\right)-1\right)+1}$, and $\beta = \mathcal{L}(\Phi_g)$ for $\theta \in \Theta$, $t \in [0,T]$ in the notation of Lemma 3.8) show that for all $\theta \in \Theta$, $t \in [0,T]$ it holds that

$$g\big(\cdot + W^\theta_T(\omega) - W^\theta_t(\omega)\big) \in \mathcal{R}\bigg(\Big\{\Phi \in \mathcal{N}\colon \mathcal{L}(\Phi) = \mathfrak{n}_{(n+1)\left(\dim\left(\mathcal{L}(\Phi_f)\right)-1\right)+1} \odot \mathcal{L}(\Phi_g)\Big\}\bigg). \tag{85}$$

Next, the induction hypothesis implies for all $\theta \in \Theta$, $t \in [0,T]$, $l \in [0,n] \cap \mathbb{N}_0$ that

$$U^\theta_{l,M}(t,\cdot,\omega) = \mathcal{R}(\Phi^\theta_{l,t}) \quad \text{and} \quad \mathcal{L}\left(\Phi^\theta_{l,t}\right) = \mathcal{L}\left(\Phi^0_{l,0}\right). \tag{86}$$

This and Lemma 3.7 (applied with

$$\begin{aligned} &d = d, \quad m = 1, \quad a = 0, \quad b = W^\theta_{\mathcal{U}^\theta_t(\omega)}(\omega) - W^\theta_t(\omega), \quad \text{and} \\ &\Psi = \Phi^\eta_{l,\mathcal{U}^\theta_t(\omega)} \quad \text{for} \quad \theta,\eta \in \Theta, \quad t \in [0,T], \quad l \in [0,n] \cap \mathbb{N}_0 \end{aligned} \tag{87}$$

in the notation of Lemma 3.7) imply that for all $\theta,\eta \in \Theta$, $t \in [0,T]$, $l \in [0,n] \cap \mathbb{N}_0$ it holds that

$$\begin{aligned} &U^\eta_{l,M}\left(\mathcal{U}^\theta_t(\omega), \cdot + W^\theta_{\mathcal{U}^\theta_t(\omega)}(\omega) - W^\theta_t(\omega), \omega\right) \\ &= \left(\mathcal{R}\big(\Phi^\eta_{l,\mathcal{U}^\theta_t(\omega)}\big)\right)\left(\cdot + W^\theta_{\mathcal{U}^\theta_t(\omega)}(\omega) - W^\theta_t(\omega)\right) \\ &\in \mathcal{R}\bigg(\Big\{\Phi \in \mathcal{N}\colon \mathcal{L}(\Phi) = \mathcal{L}\left(\Phi^\eta_{l,\mathcal{U}^\theta_t(\omega)}\right)\Big\}\bigg) = \mathcal{R}\bigg(\Big\{\Phi \in \mathcal{N}\colon \mathcal{L}(\Phi) = \mathcal{L}\left(\Phi^0_{l,0}\right)\Big\}\bigg). \end{aligned} \tag{88}$$

Moreover, Lemma 3.6 (applied with $H = (n-l)\left(\dim\left(\mathcal{L}(\Phi_f)\right) - 1\right) - 1$ for $l \in [0,n-1] \cap \mathbb{N}_0$ in the notation of Lemma 3.6) ensures for all $l \in [0,n-1] \cap \mathbb{N}_0$ that

$$\mathrm{Id}_{\mathbb{R}} \in \mathcal{R}\left(\Big\{\Phi \in \mathcal{N}\colon \mathcal{L}(\Phi) = \mathfrak{n}_{(n-l)\left(\dim\left(\mathcal{L}(\Phi_f)\right)-1\right)+1}\Big\}\right). \tag{89}$$

This, (88), and Lemma 3.8 (applied with

$$\begin{aligned} &d_1 = d, \quad d_2 = 1, \quad d_3 = 1, \quad f = \mathrm{Id}_{\mathbb{R}}, \quad \alpha = \mathfrak{n}_{(n-l)\left(\dim\left(\mathcal{L}(\Phi_f)\right)-1\right)+1}, \\ &\beta = \mathcal{L}\left(\Phi^0_{l,0}\right), \quad \text{and} \quad u = U^\eta_{l,M}\left(\mathcal{U}^\theta_t(\omega), \cdot + W^\theta_{\mathcal{U}^\theta_t(\omega)}(\omega) - W^\theta_t(\omega), \omega\right) \\ &\qquad \text{for} \quad \eta,\theta \in \Theta, \quad t \in [0,T], \quad l \in [0,n-1] \cap \mathbb{N}_0 \end{aligned} \tag{90}$$

in the notation of Lemma 3.8) prove for all $\eta,\theta \in \Theta$, $t \in [0,T]$, $l \in [0,n-1] \cap \mathbb{N}_0$ that

$$\begin{aligned} &U^\eta_{l,M}\left(\mathcal{U}^\theta_t(\omega), \cdot + W^\theta_{\mathcal{U}^\theta_t(\omega)}(\omega) - W^\theta_t(\omega), \omega\right) \\ &\in \mathcal{R}\bigg(\Big\{\Phi \in \mathcal{N}\colon \mathcal{L}(\Phi) = \mathfrak{n}_{(n-l)\left(\dim\left(\mathcal{L}(\Phi_f)\right)-1\right)+1} \odot \mathcal{L}(\Phi^0_{l,0})\Big\}\bigg). \end{aligned} \tag{91}$$

This and Lemma 3.8 (applied with

$$\begin{aligned} &d_1 = d, \quad d_2 = 1, \quad d_3 = 1, \quad f = f, \quad \alpha = \mathcal{L}(\Phi_f), \\ &\beta = \mathfrak{n}_{(n-l)\left(\dim\left(\mathcal{L}(\Phi_f)\right)-1\right)+1} \odot \mathcal{L}(\Phi^0_{l,0}), \quad \text{and} \quad u = U^\eta_{l,M}\left(\mathcal{U}^\theta_t(\omega), \cdot + W^\theta_{\mathcal{U}^\theta_t(\omega)}(\omega) - W^\theta_t(\omega), \omega\right) \\ &\qquad \text{for} \quad \eta,\theta \in \Theta, \quad t \in [0,T], \quad l \in [0,n-1] \cap \mathbb{N}_0 \end{aligned} \tag{92}$$

in the notation of Lemma 3.8) assure for all $\eta,\theta \in \Theta$, $t \in [0,T]$, $l \in [0,n-1] \cap \mathbb{N}_0$ that

$$\begin{aligned} &\left(f \circ U^\eta_{l,M}\right)\left(\mathcal{U}^\theta_t(\omega), \cdot + W^\theta_{\mathcal{U}^\theta_t(\omega)}(\omega) - W^\theta_t(\omega), \omega\right) \\ &\in \mathcal{R}\bigg(\Big\{\Phi \in \mathcal{N}\colon \mathcal{L}(\Phi) = \mathcal{L}(\Phi_f) \odot \mathfrak{n}_{(n-l)\left(\dim\left(\mathcal{L}(\Phi_f)\right)-1\right)+1} \odot \mathcal{L}(\Phi^0_{l,0})\Big\}\bigg). \end{aligned} \tag{93}$$

Next, (88) (applied with $l = n$) and Lemma 3.8 (applied with

$$\begin{aligned} &d_1 = d, \quad d_2 = 1, \quad d_3 = 1, \quad f = f, \quad \alpha = \mathcal{L}(\Phi_f), \quad \beta = \mathcal{L}\left(\Phi^0_{n,0}\right), \quad \text{and} \\ &u = \left(U^\eta_{n,M}\right)\left(\mathcal{U}^\theta_t(\omega), \cdot + W^\theta_{\mathcal{U}^\theta_t(\omega)}(\omega) - W^\theta_t(\omega), \omega\right) \quad \text{for} \quad \eta, \theta \in \Theta, \quad t \in [0,T] \end{aligned} \tag{94}$$

in the notation of Lemma 3.8) prove for all $\eta, \theta \in \Theta$, $t \in [0,T]$ that

$$\begin{aligned} &\left(f \circ U^\eta_{n,M}\right)\left(\mathcal{U}^\theta_t(\omega), \cdot + W^\theta_{\mathcal{U}^\theta_t(\omega)}(\omega) - W^\theta_t(\omega), \omega\right) \\ &\in \mathcal{R}\left(\left\{\Phi \in \mathcal{N} \colon \mathcal{L}(\Phi) = \mathcal{L}(\Phi_f) \odot \mathcal{L}(\Phi^0_{n,0})\right\}\right). \end{aligned} \tag{95}$$

Furthermore, the definition of $\odot$ in (37) and the fact that

$$\forall l \in [0,n] \cap \mathbb{N}_0 \colon \dim\left(\mathcal{L}(\Phi^0_{l,0})\right) = l\left(\dim\left(\mathcal{L}\left(\Phi_f\right)\right) - 1\right) + \dim\left(\mathcal{L}\left(\Phi_g\right)\right) \tag{96}$$

in the induction hypothesis imply that

$$\begin{aligned} &\dim\left(\mathfrak{n}_{(n+1)\left(\dim\left(\mathcal{L}(\Phi_f)\right)-1\right)+1} \odot \mathcal{L}(\Phi_g)\right) \\ &= \left[(n+1)\Big(\dim\left(\mathcal{L}(\Phi_f)\right) - 1\Big) + 1\right] + \dim\left(\mathcal{L}\left(\Phi_g\right)\right) - 1 \\ &= (n+1)\Big(\dim\left(\mathcal{L}(\Phi_f)\right) - 1\Big) + \dim\left(\mathcal{L}\left(\Phi_g\right)\right), \end{aligned} \tag{97}$$

that

$$\begin{aligned} &\dim\left(\mathcal{L}(\Phi_f) \odot \mathcal{L}(\Phi^0_{n,0})\right) = \dim\left(\mathcal{L}(\Phi_f)\right) + \dim\left(\mathcal{L}(\Phi^0_{n,0})\right) - 1 \\ &= \dim\left(\mathcal{L}(\Phi_f)\right) + \left[n\Big(\dim\left(\mathcal{L}\left(\Phi_f\right)\right) - 1\Big) + \dim\left(\mathcal{L}\left(\Phi_g\right)\right)\right] - 1 \\ &= (n+1)\Big(\dim\left(\mathcal{L}(\Phi_f)\right) - 1\Big) + \dim\left(\mathcal{L}\left(\Phi_g\right)\right), \end{aligned} \tag{98}$$

and for all $l \in [0, n-1] \cap \mathbb{N}_0$ that

$$\begin{aligned} &\dim\left(\mathcal{L}(\Phi_f) \odot \mathfrak{n}_{(n-l)\left(\dim\left(\mathcal{L}(\Phi_f)\right)-1\right)+1} \odot \mathcal{L}(\Phi^0_{l,0})\right) \\ &= \dim\left(\mathcal{L}(\Phi_f)\right) + \dim\left(\mathfrak{n}_{(n-l)\left(\dim\left(\mathcal{L}(\Phi_f)\right)-1\right)+1}\right) + \dim\left(\mathcal{L}(\Phi^0_{l,0})\right) - 2 \\ &= \dim\left(\mathcal{L}(\Phi_f)\right) + \left[(n-l)\left(\dim\left(\mathcal{L}(\Phi_f)\right) - 1\right) + 1\right] \\ &\qquad + \left[l\left(\dim\left(\mathcal{L}\left(\Phi_f\right)\right) - 1\right) + \dim\left(\mathcal{L}\left(\Phi_g\right)\right)\right] - 2 \\ &= \dim\left(\mathcal{L}(\Phi_f)\right) + n\left(\dim\left(\mathcal{L}(\Phi_f)\right) - 1\right) + \dim\left(\mathcal{L}\left(\Phi_g\right)\right) - 1 \\ &= (n+1)\left(\dim\left(\mathcal{L}(\Phi_f)\right) - 1\right) + \dim\left(\mathcal{L}\left(\Phi_g\right)\right). \end{aligned} \tag{99}$$

This shows, roughly speaking, that the functions in (85), (95), and (93) can be represented by networks with the same depth (i.e. number of layers): $(n+1)(\dim\left(\mathcal{L}(\Phi_f)\right) - 1) + \dim\left(\mathcal{L}\left(\Phi_g\right)\right)$. Hence, Lemma 3.9 and (79) imply that there exists a family $(\Phi^\theta_{n+1,t})_{\theta \in \Theta, t \in [0,T]} \subseteq \mathcal{N}$ such that for all $\theta \in \Theta$, $t \in [0, T]$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
&\big(\mathcal{R}(\Phi^{\theta}_{n+1,t})\big)(x) \\
&= \frac{1}{M^{n+1}} \sum_{i=1}^{M^{n+1}} g\big(x + W^{(\theta,0,-i)}_T(\omega) - W^{(\theta,0,-i)}_t(\omega)\big) \\
&\quad + \frac{(T-t)}{M} \sum_{i=1}^{M} \Big(f \circ U^{(\theta,n,i)}_{n,M}\Big) \Big(\mathcal{U}^{(\theta,n,i)}_t(\omega), x + W^{(\theta,n,i)}_{\mathcal{U}^{(\theta,n,i)}_t(\omega)}(\omega) - W^{(\theta,n,i)}_t(\omega), \omega\Big) \\
&\quad + \sum_{l=0}^{n-1} \frac{(T-t)}{M^{n+1-l}} \sum_{i=1}^{M^{n+1-l}} \Big(f \circ U^{(\theta,l,i)}_{l,M}\Big) \Big(\mathcal{U}^{(\theta,l,i)}_t(\omega), x + W^{(\theta,l,i)}_{\mathcal{U}^{(\theta,l,i)}_t(\omega)}(\omega) - W^{(\theta,l,i)}_t(\omega), \omega\Big) \\
&\quad - \sum_{l=1}^{n} \frac{(T-t)}{M^{n+1-l}} \sum_{i=1}^{M^{n+1-l}} \Big(f \circ U^{(\theta,-l,i)}_{l-1,M}\Big) \Big(\mathcal{U}^{(\theta,l,i)}_t(\omega), x + W^{(\theta,l,i)}_{\mathcal{U}^{(\theta,l,i)}_t(\omega)}(\omega) - W^{(\theta,l,i)}_t(\omega), \omega\Big) \\
&= U^{\theta}_{n+1,M}(t,x,\omega),
\end{aligned} \tag{100}$$

that

$$\dim\big(\mathcal{L}(\Phi^{\theta}_{n+1,t})\big) = (n+1)(\dim(\mathcal{L}(\Phi_f)) - 1) + \dim(\mathcal{L}(\Phi_g)), \tag{101}$$

and that

$$\begin{aligned}
\mathcal{L}(\Phi^{\theta}_{n+1,t}) = &\left(\mathop{\boxplus}_{i=1}^{M^{n+1}} \Big[\mathfrak{n}_{(n+1)(\dim(\mathcal{L}(\Phi_f))-1)+1} \odot \mathcal{L}(\Phi_g)\Big]\right) \boxplus \left(\mathop{\boxplus}_{i=1}^{M} \big(\mathcal{L}(\Phi_f) \odot \mathcal{L}(\Phi^0_{n,0})\big)\right) \\
&\boxplus \left(\mathop{\boxplus}_{l=0}^{n-1} \mathop{\boxplus}_{i=1}^{M^{n+1-l}} \Bigg[\Big(\mathcal{L}(\Phi_f) \odot \mathfrak{n}_{(n-l)(\dim(\mathcal{L}(\Phi_f))-1)+1} \odot \mathcal{L}(\Phi^0_{l,0})\Big)\right) \\
&\boxplus \left(\mathop{\boxplus}_{l=1}^{n} \mathop{\boxplus}_{i=1}^{M^{n+1-l}} \Big(\mathcal{L}(\Phi_f) \odot \mathfrak{n}_{(n-l+1)(\dim(\mathcal{L}(\Phi_f))-1)+1} \odot \mathcal{L}(\Phi^0_{l-1,0})\Big)\Bigg]\right).
\end{aligned} \tag{102}$$

This shows for all $t_1, t_2 \in [0, T]$, $\theta_1, \theta_2 \in \Theta$ that

$$\mathcal{L}\left(\Phi^{\theta_1}_{n+1,t_1}\right) = \mathcal{L}\left(\Phi^{\theta_2}_{n+1,t_2}\right). \tag{103}$$

Furthermore, (102), the triangle inequality (see Lemma 3.5), and the fact that

$$\forall l \in [0, n] \cap \mathbb{N}_0 \colon \ \big\|\mathcal{L}(\Phi^0_{l,0})\big\|_\infty \le c(3M)^l \tag{104}$$

in the induction hypothesis show for all $\theta \in \Theta$, $t \in [0, T]$ that

$$\begin{aligned}
\big\|\mathcal{L}(\Phi^{\theta}_{n+1,t})\big\|_\infty \le &\sum_{i=1}^{M^{n+1}} \Big\|\mathfrak{n}_{(n+1)(\dim(\mathcal{L}(\Phi_f))-1)+1} \odot \mathcal{L}(\Phi_g)\Big\|_\infty + \sum_{i=1}^{M} \big\|\mathcal{L}(\Phi_f) \odot \mathcal{L}(\Phi^0_{n,0})\big\|_\infty \\
&+ \sum_{l=0}^{n-1} \sum_{i=1}^{M^{n+1-l}} \Big\|\mathcal{L}(\Phi_f) \odot \mathfrak{n}_{(n-l)(\dim(\mathcal{L}(\Phi_f))-1)+1} \odot \mathcal{L}(\Phi^0_{l,0})\Big\|_\infty \\
&+ \sum_{l=1}^{n} \sum_{i=1}^{M^{n+1-l}} \Big\|\mathcal{L}(\Phi_f) \odot \mathfrak{n}_{(n-l+1)(\dim(\mathcal{L}(\Phi_f))-1)+1} \odot \mathcal{L}(\Phi^0_{l-1,0})\Big\|_\infty.
\end{aligned} \tag{105}$$

Note that for all $H_1, H_2, \alpha_0, \dots, \alpha_{H_1+1}, \beta_0, \dots, \beta_{H_2+1} \in \mathbb{N}$, $\alpha, \beta \in \mathcal{D}$ with $\alpha = (\alpha_0, \dots, \alpha_{H_1+1})$, $\beta = (\beta_0, \dots, \beta_{H_2+1})$, $\alpha_0 = \beta_{H_2+1} = 1$ it holds that $\|\alpha \odot \beta\|_\infty \le \max\{\|\alpha\|_\infty, \|\beta\|_\infty, 2\}$ (see (37)).

This, (105), the fact that $\forall H \in \mathbb{N}\colon \|\mathfrak{n}_{H+2}\|_\infty = 2$ (see (41)), (78), and (104) prove that

$$\begin{aligned}
&\left\|\mathcal{L}(\Phi^\theta_{n+1,t})\right\|_\infty \\
&\leq \left[\sum_{i=1}^{M^{n+1}} c\right] + \left[\sum_{i=1}^{M} c(3M)^n\right] + \left[\sum_{l=0}^{n-1}\sum_{i=1}^{M^{n+1-l}} c(3M)^l\right] + \left[\sum_{l=1}^{n}\sum_{i=1}^{M^{n+1-l}} c(3M)^{l-1}\right] \\
&= M^{n+1}c + Mc(3M)^n + \left[\sum_{l=0}^{n-1} M^{n+1-l}c(3M)^l\right] + \left[\sum_{l=1}^{n} M^{n+1-l}c(3M)^{l-1}\right] \\
&= M^{n+1}c\left[1 + 3^n + \sum_{l=0}^{n-1} 3^l + \sum_{l=1}^{n} 3^{l-1}\right] = M^{n+1}c\left[1 + \sum_{l=0}^{n} 3^l + \sum_{l=1}^{n} 3^{l-1}\right] \\
&\leq cM^{n+1}\left[1 + 2\sum_{l=0}^{n} 3^l\right] = cM^{n+1}\left[1 + 2\frac{3^{n+1}-1}{3-1}\right] = c(3M)^{n+1}.
\end{aligned} \tag{106}$$

Combining (100), (101), (103), and (106) completes the induction step. Induction hence establishes Item (i)–Item (iii). The proof of Lemma 3.10 is thus completed. ☐

4 Main result

Theorem 4.1. *Let $T, L \in (0,\infty)$, $B \in [2,\infty)$, $p, \tilde{p} \in \mathbb{N}$, $q \in \mathbb{N} \cap [2,\infty)$, $\alpha, \beta \in [0,\infty)$, $f \in C(\mathbb{R},\mathbb{R})$, for every $d \in \mathbb{N}$ let $g_d \in C(\mathbb{R}^d,\mathbb{R})$, for every $d \in \mathbb{N}$ let $\nu_d\colon \mathcal{B}(\mathbb{R}^d) \to [0,1]$ be a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, for every $d \in \mathbb{N}$ let $\mathbf{A}_d\colon \mathbb{R}^d \to \mathbb{R}^d$ be the function such that for all $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$ it holds that*

$$\mathbf{A}_d(x) = (\max\{x_1,0\},\ldots,\max\{x_d,0\}), \tag{107}$$

let $\mathcal{N}$ and $\mathcal{D}$ be the sets which satisfy

$$\mathcal{N} = \bigcup_{H\in\mathbb{N}} \bigcup_{(k_0,k_1,\ldots,k_H,k_{H+1})\in\mathbb{N}^{H+2}} \prod_{n=1}^{H+1} \left(\mathbb{R}^{k_n\times k_{n-1}} \times \mathbb{R}^{k_n}\right) \quad \text{and} \quad \mathcal{D} = \bigcup_{H\in\mathbb{N}} \mathbb{N}^{H+2}, \tag{108}$$

let

$$\mathcal{P}\colon \mathcal{N} \to \mathbb{N}, \quad \mathcal{L}\colon \mathcal{N} \to \mathcal{D}, \quad \text{and} \quad \mathcal{R}\colon \mathcal{N} \to \bigcup_{k,l\in\mathbb{N}} C(\mathbb{R}^k,\mathbb{R}^l) \tag{109}$$

be the functions such that for all $H \in \mathbb{N}$, $k_0, k_1, \ldots, k_H, k_{H+1} \in \mathbb{N}$,

$$\Phi = ((W_1,B_1),\ldots,(W_{H+1},B_{H+1})) \in \prod_{n=1}^{H+1} \left(\mathbb{R}^{k_n\times k_{n-1}} \times \mathbb{R}^{k_n}\right), \tag{110}$$

$x_0 \in \mathbb{R}^{k_0}, \ldots, x_H \in \mathbb{R}^{k_H}$ with $\forall n \in \mathbb{N} \cap [1,H]\colon x_n = \mathbf{A}_{k_n}(W_n x_{n-1} + B_n)$ it holds that

$$\begin{gathered}
\mathcal{P}(\Phi) = \sum_{n=1}^{H+1} k_n(k_{n-1}+1), \quad \mathcal{L}(\Phi) = (k_0,k_1,\ldots,k_H,k_{H+1}), \\
\mathcal{R}(\Phi) \in C(\mathbb{R}^{k_0},\mathbb{R}^{k_{H+1}}), \quad \text{and} \quad (\mathcal{R}(\Phi))(x_0) = W_{H+1}x_H + B_{H+1},
\end{gathered} \tag{111}$$

for every $\varepsilon \in (0,1]$, $d \in \mathbb{N}$ let $\Phi^f_\varepsilon, \Phi^{g_d}_\varepsilon \in \mathcal{N}$, assume for all $d \in \mathbb{N}$, $v,w \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$ that $\mathcal{R}(\Phi^f_\varepsilon) \in C(\mathbb{R},\mathbb{R})$, $\mathcal{R}(\Phi^{g_d}_\varepsilon) \in C(\mathbb{R}^d,\mathbb{R})$, $\left|(\mathcal{R}(\Phi^f_\varepsilon))(w) - (\mathcal{R}(\Phi^f_\varepsilon))(v)\right| \leq L\,|w-v|$, $\left|(\mathcal{R}(\Phi^f_\varepsilon))(0)\right| \leq B$, $|(\mathcal{R}(\Phi^{g_d}_\varepsilon))(x)| \leq Bd^p(1+\|x\|)^p$, $\left|f(v) - (\mathcal{R}(\Phi^f_\varepsilon))(v)\right| \leq \varepsilon B(1+|v|^q)$, $|g_d(x) - (\mathcal{R}(\Phi^{g_d}_\varepsilon))(x)| \leq \varepsilon Bd^p(1+\|x\|)^{pq}$, $\max\left\{\left\|\mathcal{L}(\Phi^f_\varepsilon)\right\|_\infty, \|\mathcal{L}(\Phi^{g_d}_\varepsilon)\|_\infty\right\} \leq d^p\varepsilon^{-\alpha}B$, $\max\left\{\dim\left(\mathcal{L}\left(\Phi^f_\varepsilon\right)\right), \dim\left(\mathcal{L}\left(\Phi^{g_d}_\varepsilon\right)\right)\right\} \leq d^p\varepsilon^{-\beta}B$, and $\left(\int_{\mathbb{R}^d} \|y\|^{2pq}\,\nu_d(dy)\right)^{1/(2pq)} \leq Bd^{\tilde{p}}$. Then *i) for every $d \in \mathbb{N}$ there exists a unique continuous function $u_d\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ such that for every $x \in \mathbb{R}^d$, for every $s \in [0,T]$, for every probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and for every standard Brownian motion $\mathbf{W}\colon [0,T] \times \Omega \to \mathbb{R}^d$ with continuous sample paths it holds that $\sup_{t\in[0,T]} \sup_{y\in\mathbb{R}^d} \left(\frac{|u_d(t,y)|}{1+\|y\|^p}\right) < \infty$ and*

$$u_d(s,x) = \mathbb{E}\left[g_d\left(x + \mathbf{W}_{T-s}\right) + \int_s^T f\left(u_d\left(t, x + \mathbf{W}_{t-s}\right)\right) dt\right] \tag{112}$$

and

ii) there exist $(\Psi_{d,\varepsilon})_{d\in\mathbb{N},\varepsilon\in(0,1]} \subseteq \mathcal{N}$, $\eta \in (0,\infty)$, $C\colon (0,1] \to (0,\infty)$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\gamma \in (0,1]$ it holds that

$$\mathcal{P}(\Psi_{d,\varepsilon}) \leq C(\gamma) d^\eta \varepsilon^{-(4+2\alpha+\beta+\gamma)}, \tag{113}$$

$\mathcal{R}(\Psi_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R})$, and

$$\left(\int_{\mathbb{R}^d} \left|u_d(0,x) - (\mathcal{R}(\Psi_{d,\varepsilon}))(x)\right|^2 \nu_d(dx)\right)^{1/2} \leq \varepsilon. \tag{114}$$

Proof of Theorem 4.1. First note that the triangle inequality, the fact that $\forall v, w \in \mathbb{R}, \varepsilon \in (0,1]\colon \left|(\mathcal{R}(\Phi^f_\varepsilon))(w) - (\mathcal{R}(\Phi^f_\varepsilon))(v)\right| \leq L\left|w - v\right|$, and the fact that $\forall v \in \mathbb{R}, \varepsilon \in (0,1]\colon \left|f(v) - (\mathcal{R}(\Phi^f_\varepsilon))(v)\right| \leq \varepsilon B(1 + |v|^q)$ imply for all $v, w \in \mathbb{R}, \varepsilon \in (0,1]$ that

$$\begin{aligned} |f(w) - f(v)| &\leq \left|f(w) - (\mathcal{R}(\Phi^f_\varepsilon))(w)\right| + \left|(\mathcal{R}(\Phi^f_\varepsilon))(w) - (\mathcal{R}(\Phi^f_\varepsilon))(v)\right| + \left|(\mathcal{R}(\Phi^f_\varepsilon))(v) - f(v)\right| \\ &\leq \varepsilon B(1 + |w|^q) + L\left|w - v\right| + \varepsilon B(1 + |v|^q). \end{aligned} \tag{115}$$

This proves that for all $v, w \in \mathbb{R}$ it holds that

$$|f(w) - f(v)| \leq L\left|w - v\right|. \tag{116}$$

The triangle inequality, the fact that $\forall \varepsilon \in (0,1]\colon \left|(\mathcal{R}(\Phi^f_\varepsilon))(0)\right| \leq B$, and the fact that $\forall \varepsilon \in (0,1]\colon \left|f(0) - (\mathcal{R}(\Phi^f_\varepsilon))(0)\right| \leq \varepsilon B$ imply for all $\varepsilon \in (0,1]$ that

$$|f(0)| \leq \left|f(0) - (\mathcal{R}(\Phi^f_\varepsilon))(0)\right| + \left|(\mathcal{R}(\Phi^f_\varepsilon))(0)\right| \leq \varepsilon B + B. \tag{117}$$

This proves that

$$|f(0)| \leq B. \tag{118}$$

The triangle inequality, the fact that $\forall d \in \mathbb{N}, x \in \mathbb{R}^d, \varepsilon \in (0,1]\colon |(\mathcal{R}(\Phi^{g_d}_\varepsilon))(x)| \leq Bd^p(1+\|x\|)^p$, and the fact that $\forall d \in \mathbb{N}, x \in \mathbb{R}^d, \varepsilon \in (0,1]\colon |g_d(x) - (\mathcal{R}(\Phi^{g_d}_\varepsilon))(x)| \leq \varepsilon Bd^p(1 + \|x\|)^{pq}$ imply for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$ that

$$|g_d(x)| \leq |g_d(x) - (\mathcal{R}(\Phi^{g_d}_\varepsilon))(x)| + |(\mathcal{R}(\Phi^{g_d}_\varepsilon))(x)| \leq \varepsilon Bd^p(1 + \|x\|)^{pq} + Bd^p(1 + \|x\|)^p. \tag{119}$$

This proves for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$ that

$$|g_d(x)| \leq Bd^p(1 + \|x\|)^p. \tag{120}$$

Item (i) follows from Corollary 3.11 in [14] together with (116) and (120). It thus remains to prove Item (ii). To this end let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, for every $d \in \mathbb{N}$ let $\mathbf{W}^d\colon [0,T] \times \Omega \to \mathbb{R}^d$ be a standard Brownian motion with continuous sample paths, let $\Theta = \bigcup_{n\in\mathbb{N}} \mathbb{Z}^n$, let $\mathfrak{u}^\theta\colon \Omega \to [0,1]$, $\theta \in \Theta$, be independent random variables which are uniformly distributed on $[0,1]$, let $\mathcal{U}^\theta\colon [0,T] \times \Omega \to [0,T]$, $\theta \in \Theta$, satisfy for all $t \in [0,T]$, $\theta \in \Theta$ that

$\mathcal{U}_t^\theta = t + (T-t)\mathfrak{u}^\theta$, for every $d \in \mathbb{N}$ let $W^{\theta,d}\colon [0,T]\times\Omega \to \mathbb{R}^d$, $\theta\in\Theta$, be independent standard Brownian motions with continuous sample paths, assume for every $d \in \mathbb{N}$ that $(\mathfrak{u}^\theta)_{\theta\in\Theta}$ and $(W^{\theta,d})_{\theta\in\Theta}$ are independent, and let $U^\theta_{n,M,d,\delta}\colon [0,T]\times\mathbb{R}^d\times\Omega\to\mathbb{R}$, $n,M\in\mathbb{Z}$, $d\in\mathbb{N}$, $\delta\in(0,1]$, $\theta\in\Theta$, be functions such that for all $d,n,M\in\mathbb{N}$, $\delta\in(0,1]$, $\theta\in\Theta$, $t\in[0,T]$, $x\in\mathbb{R}^d$ it holds that $U^\theta_{-1,M,d,\delta}(t,x) = U^\theta_{0,M,d,\delta}(t,x) = 0$ and

$$
\begin{aligned}
U^\theta_{n,M,d,\delta}(t,x) =& \frac{1}{M^n}\sum_{i=1}^{M^n}(\mathcal{R}(\Phi^{g_d}_\delta))\big(x + W_T^{(\theta,0,-i),d} - W_t^{(\theta,0,-i),d}\big) \\
&+ \sum_{l=0}^{n-1}\frac{(T-t)}{M^{n-l}}\Bigg[\sum_{i=1}^{M^{n-l}}\big((\mathcal{R}(\Phi^f_\delta))\circ U^{(\theta,l,i)}_{l,M,d,\delta} - \mathbb{1}_{\mathbb{N}}(l)(\mathcal{R}(\Phi^f_\delta))\circ U^{(\theta,-l,i)}_{l-1,M,d,\delta}\big) \\
&\qquad\qquad \left(\mathcal{U}_t^{(\theta,l,i)}, x + W^{(\theta,l,i),d}_{\mathcal{U}_t^{(\theta,l,i)}} - W_t^{(\theta,l,i),d}\right)\Bigg],
\end{aligned}
\tag{121}
$$

let $c_d\in[1,\infty)$, $d\in\mathbb{N}$, be the real numbers that satisfy for all $d\in\mathbb{N}$ that

$$
c_d = \big(e^{LT}(T+1)\big)^{q+1}\left((Bd^p)^q+1\right)\left[1 + \left(\int_{\mathbb{R}^d}\|x\|^{2pq}\,\nu_d(dx)\right)^{1/(2pq)} + \left(\mathbb{E}\Big[\big\|\mathbf{W}^d_T\big\|^{pq}\Big]\right)^{1/(pq)}\right]^{pq},
\tag{122}
$$

let $k_{d,\varepsilon}\in\mathbb{N}$, $d\in\mathbb{N}$, $\varepsilon\in(0,1]$, be the natural numbers that satisfy for all $d\in\mathbb{N}$, $\varepsilon\in(0,1]$ that

$$
k_{d,\varepsilon} = \max\left\{\big\|\mathcal{L}(\Phi^f_\varepsilon)\big\|_\infty, \|\mathcal{L}(\Phi^{g_d}_\varepsilon)\|_\infty, 2\right\},
\tag{123}
$$

let $\tilde{C}\colon (0,\infty)\to(0,\infty]$ be the function that satisfies for all $\gamma\in(0,\infty)$ that

$$
\tilde{C}(\gamma) = \sup_{n\in\mathbb{N}\cap[2,\infty)}\left[n(3n)^{2n}\left(\frac{\sqrt{e}(1+2LT)}{\sqrt{n-1}}\right)^{(n-1)(4+\gamma)}\right],
\tag{124}
$$

let $N_{d,\varepsilon}\in\mathbb{N}$, $d\in\mathbb{N}$, $\varepsilon\in(0,1]$, be the natural numbers that satisfy for all $d\in\mathbb{N}$, $\varepsilon\in(0,1]$ that

$$
N_{d,\varepsilon} = \min\left\{n\in\mathbb{N}\cap[2,\infty)\colon c_d\left(\frac{\sqrt{e}(1+2LT)}{\sqrt{n}}\right)^n \le \frac{\varepsilon}{2}\right\},
\tag{125}
$$

and let $\delta_{d,\varepsilon}\in(0,1]$, $d\in\mathbb{N}$, $\varepsilon\in(0,1]$, be the real numbers that satisfy for all $d\in\mathbb{N}$, $\varepsilon\in(0,1]$ that $\delta_{d,\varepsilon} = \frac{\epsilon}{4Bd^pc_d}$.

Note that for all $d\in\mathbb{N}$ the random variable $\big\|\mathbf{W}^d_T/\sqrt{T}\big\|^2$ is chi-squared distributed with d degrees of freedom. This and Jensen's inequality imply that for all $d\in\mathbb{N}$ it holds that

$$
\left(\mathbb{E}\big[\|\mathbf{W}^d_T\|^{pq}\big]\right)^2 \le \mathbb{E}\Big[\big\|\mathbf{W}^d_T\big\|^{2pq}\Big] = (2T)^{pq}\frac{\Gamma\left(\frac{d}{2}+pq\right)}{\Gamma\left(\frac{d}{2}\right)} = (2T)^{pq}\prod_{k=0}^{pq-1}\left(\frac{d}{2}+k\right).
\tag{126}
$$

This implies for all $d\in\mathbb{N}$ that

$$
\left(\mathbb{E}\Big[\big\|\mathbf{W}^d_T\big\|^{pq}\Big]\right)^{1/(pq)} = \left(\mathbb{E}\Big[\big\|\mathbf{W}^d_T\big\|^{pq}\Big]\right)^{2/(2pq)} \le \sqrt{2T}\left(\prod_{k=0}^{pq-1}\left(\frac{d}{2}+k\right)\right)^{1/(2pq)} \le \sqrt{2T\left(\frac{d}{2}+pq-1\right)}.
\tag{127}
$$

This together with the fact that $\forall d\in\mathbb{N}\colon \left(\int_{\mathbb{R}^d}\|x\|^{2pq}\,\nu_d(dx)\right)^{1/(2pq)} \le Bd^{\tilde{p}}$ implies that there exist $\bar{C}\in(0,\infty)$ such that for all $d\in\mathbb{N}$ it holds that

$$
c_d \le \bar{C}d^{pq}\left(\frac{1+d^{\tilde{p}}+\sqrt{d}}{3}\right)^{pq} \le \bar{C}d^{(\tilde{p}+1)pq}.
\tag{128}
$$

Next note that for all $\gamma \in (0, \infty)$ it holds that

$$\begin{aligned}
\tilde{C}(\gamma) &= \sup_{n \in \mathbb{N} \cap [2,\infty)} \left[n(3n)^{2n} \left(\frac{\sqrt{e}(1+2LT)}{\sqrt{n-1}} \right)^{(n-1)(4+\gamma)} \right] \\
&= \sup_{n \in \mathbb{N} \cap [2,\infty)} \left[(\sqrt{e}(1+2LT))^{(n-1)(4+\gamma)} n^3 3^{2n} (n-1)^{-(n-1)\frac{\gamma}{2}} \left(\frac{n}{n-1} \right)^{2(n-1)} \right] \\
&\leq \left[\sup_{n \in \mathbb{N} \cap [2,\infty)} \left[(\sqrt{e}(1+2LT))^{(n-1)(4+\gamma)} n^3 3^{2n} (n-1)^{-(n-1)\frac{\gamma}{2}} \right] \right] \left[\sup_{n \in \mathbb{N} \cap [2,\infty)} \left(\frac{n}{n-1} \right)^{2(n-1)} \right] \\
&< \infty.
\end{aligned} \tag{129}$$

The fact that for all $d \in \mathbb{N}$, $v \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$ it holds that $\left| f(v) - (\mathcal{R}(\Phi^f_\varepsilon))(v) \right| \leq \varepsilon B(1+|v|^q)$ and $|g_d(x) - (\mathcal{R}(\Phi^{g_d}_\varepsilon))(x)| \leq \varepsilon B d^p (1+\|x\|)^{pq}$ implies for all $d \in \mathbb{N}$, $v \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$ that

$$\begin{aligned}
\max\left\{ \left| f(v) - (\mathcal{R}(\Phi^f_\varepsilon))(v) \right|, |g_d(x) - (\mathcal{R}(\Phi^{g_d}_\varepsilon))(x)| \right\} &\leq \max\{\varepsilon B(1+|v|^q), \varepsilon B d^p (1+\|x\|)^{pq}\} \\
&\leq \varepsilon B d^p ((1+\|x\|)^{pq} + |v|^q).
\end{aligned} \tag{130}$$

This, (116), (118), (120), the fact that for all $d \in \mathbb{N}$, $w, v \in \mathbb{R}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0,1]$ it holds that $\left| (\mathcal{R}(\Phi^f_\varepsilon))(w) - (\mathcal{R}(\Phi^f_\varepsilon))(v) \right| \leq L|w-v|$, $\left| (\mathcal{R}(\Phi^f_\varepsilon))(0) \right| \leq B$, $|(\mathcal{R}(\Phi^{g_d}_\varepsilon))(x)| \leq B d^p (1+\|x\|)^p$, and Corollary 2.4 (with $f_1 = f$, $f_2 = \mathcal{R}(\Phi^f_\delta)$, $g_1 = g_d$, $g_2 = \mathcal{R}(\Phi^{g_d}_\delta)$, $L = L$, $\delta = \delta B d^p$, $B = B d^p$, $\mathbf{W} = \mathbf{W}^d$ in the notation of Corollary 2.4), imply for all $d, N, M \in \mathbb{N}$, $\delta \in (0,1]$ that

$$\begin{aligned}
&\left(\int_{\mathbb{R}^d} \mathbb{E}\left[\left| U^0_{N,M,d,\delta}(0,x) - u_d(0,x) \right|^2 \right] \nu_d(dx) \right)^{1/2} \\
&\qquad \leq \left(e^{LT}(T+1) \right)^{q+1} ((Bd^p)^q + 1) \left(\delta B d^p + \frac{e^{M/2}(1+2LT)^N}{M^{N/2}} \right) \\
&\qquad\qquad \cdot \left[\int_{\mathbb{R}^d} \left(1 + \|x\| + \left(\mathbb{E}\left[\left\| \mathbf{W}^d_T \right\|^{pq} \right] \right)^{\frac{1}{pq}} \right)^{2pq} \nu_d(dx) \right]^{1/2}.
\end{aligned} \tag{131}$$

This and the triangle inequality prove for all $d, N, M \in \mathbb{N}$, $\delta \in (0,1]$ that

$$\begin{aligned}
&\left(\int_{\mathbb{R}^d} \mathbb{E}\left[\left| U^0_{N,M,d,\delta}(0,x) - u_d(0,x) \right|^2 \right] \nu_d(dx) \right)^{1/2} \\
&\leq \left(e^{LT}(T+1) \right)^{q+1} ((Bd^p)^q + 1) \left(\delta B d^p + \frac{e^{M/2}(1+2LT)^N}{M^{N/2}} \right) \\
&\qquad \cdot \left[1 + \left(\int_{\mathbb{R}^d} \|x\|^{2pq} \nu_d(dx) \right)^{1/(2pq)} + \left(\mathbb{E}\left[\left\| \mathbf{W}^d_T \right\|^{pq} \right] \right)^{1/(pq)} \right]^{pq} \\
&= c_d \left(\delta B d^p + \frac{e^{M/2}(1+2LT)^N}{M^{N/2}} \right).
\end{aligned} \tag{132}$$

This and Fubini's theorem imply that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\begin{aligned}
&\mathbb{E}\left[\int_{\mathbb{R}^d} \left|U^0_{N_{d,\varepsilon},N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(0,x) - u_d(0,x)\right|^2 \nu_d(dx)\right] \\
&= \int_{\mathbb{R}^d} \mathbb{E}\left[\left|U^0_{N_{d,\varepsilon},N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(0,x) - u_d(0,x)\right|^2\right] \nu_d(dx) \\
&\leq \left(c_d\delta_{d,\varepsilon}Bd^p + c_d\left(\frac{\sqrt{e}(1+2LT)}{\sqrt{N_{d,\varepsilon}}}\right)^{N_{d,\varepsilon}}\right)^2 \leq \left(\frac{\varepsilon}{4}+\frac{\varepsilon}{2}\right)^2 < \varepsilon^2.
\end{aligned} \tag{133}$$

This implies that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ there exists $\omega_{d,\varepsilon} \in \Omega$ such that

$$\int_{\mathbb{R}^d} \left|U^0_{N_{d,\varepsilon},N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(0,x,\omega_{d,\varepsilon}) - u_d(0,x)\right|^2 \nu_d(dx) < \varepsilon^2. \tag{134}$$

Next, Lemma 3.10 shows that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ there exists $\Psi_{d,\varepsilon} \in \mathcal{N}$ such that for all $x \in \mathbb{R}^d$ it holds that $\mathcal{R}(\Psi_{d,\varepsilon}) \in C(\mathbb{R}^d,\mathbb{R})$, $(\mathcal{R}(\Psi_{d,\varepsilon}))(x) = U^0_{N_{d,\varepsilon},N_{d,\varepsilon},d,\delta_{d,\varepsilon}}(0,x,\omega_{d,\varepsilon})$,

$$\dim\left(\mathcal{L}\left(\Psi_{d,\varepsilon}\right)\right) = N_{d,\varepsilon}\left(\dim\left(\mathcal{L}\left(\Phi^f_{\delta_{d,\varepsilon}}\right)\right) - 1\right) + \dim\left(\mathcal{L}\left(\Phi^{g_d}_{\delta_{d,\varepsilon}}\right)\right), \tag{135}$$

and

$$\|\mathcal{L}(\Psi_{d,\varepsilon})\|_\infty \leq k_{d,\delta_{d,\varepsilon}}(3N_{d,\varepsilon})^{N_{d,\varepsilon}}. \tag{136}$$

This and (134) prove (114). Moreover, this and (111) imply for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that

$$\begin{aligned}
\mathcal{P}(\Psi_{d,\varepsilon}) &\leq \sum_{j=1}^{\dim(\mathcal{L}(\Psi_{d,\varepsilon}))} k_{d,\delta_{d,\varepsilon}}(3N_{d,\varepsilon})^{N_{d,\varepsilon}}\left(k_{d,\delta_{d,\varepsilon}}(3N_{d,\varepsilon})^{N_{d,\varepsilon}}+1\right) \\
&\leq 2\dim\left(\mathcal{L}\left(\Psi_{d,\varepsilon}\right)\right)k^2_{d,\delta_{d,\varepsilon}}(3N_{d,\varepsilon})^{2N_{d,\varepsilon}} \\
&= 2\left(N_{d,\varepsilon}\left(\dim\left(\mathcal{L}\left(\Phi^f_{\delta_{d,\varepsilon}}\right)\right)-1\right)+\dim\left(\mathcal{L}\left(\Phi^{g_d}_{\delta_{d,\varepsilon}}\right)\right)\right)k^2_{d,\delta_{d,\varepsilon}}(3N_{d,\varepsilon})^{2N_{d,\varepsilon}}.
\end{aligned} \tag{137}$$

This together with the fact that $\forall d \in \mathbb{N}, \varepsilon \in (0,1]\colon \max\left\{\left\|\mathcal{L}(\Phi^f_\varepsilon)\right\|_\infty, \|\mathcal{L}(\Phi^{g_d}_\varepsilon)\|_\infty\right\} \leq d^p\varepsilon^{-\alpha}B$ and the fact that $\forall d \in \mathbb{N}, \varepsilon \in (0,1]\colon \max\left\{\dim\left(\mathcal{L}\left(\Phi^f_\varepsilon\right)\right), \dim\left(\mathcal{L}\left(\Phi^{g_d}_\varepsilon\right)\right), 1\right\} \leq d^p\varepsilon^{-\beta}B$ implies for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ that $k_{d,\delta_{d,\varepsilon}} \leq d^p\delta^{-\alpha}_{d,\varepsilon}B$ and that

$$\begin{aligned}
\mathcal{P}(\Psi_{d,\varepsilon}) &\leq 2\left(N_{d,\varepsilon}\left(\dim\left(\mathcal{L}\left(\Phi^f_{\delta_{d,\varepsilon}}\right)\right)-1\right)+\dim\left(\mathcal{L}\left(\Phi^{g_d}_{\delta_{d,\varepsilon}}\right)\right)\right)\left(d^p\delta^{-\alpha}_{d,\varepsilon}B\right)^2(3N_{d,\varepsilon})^{2N_{d,\varepsilon}} \\
&\leq 4d^p\delta^{-\beta}_{d,\varepsilon}Bd^{2p}\delta^{-2\alpha}_{d,\varepsilon}B^2N_{d,\varepsilon}(3N_{d,\varepsilon})^{2N_{d,\varepsilon}} \\
&= 4B^3(4c_dBd^p)^{2\alpha+\beta}d^{3p}\varepsilon^{-(2\alpha+\beta)}N_{d,\varepsilon}(3N_{d,\varepsilon})^{2N_{d,\varepsilon}}.
\end{aligned} \tag{138}$$

It follows from (125) that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$ it holds that

$$\varepsilon \leq 2c_d\left(\frac{\sqrt{e}(1+2LT)}{\sqrt{N_{d,\varepsilon}-1}}\right)^{N_{d,\varepsilon}-1}. \tag{139}$$

This together with (138) implies that for all $d \in \mathbb{N}$, $\varepsilon \in (0,1]$, $\gamma \in (0,1]$ it holds that

$$\begin{aligned}
&\mathcal{P}(\Psi_{d,\varepsilon}) \leq 4B^{2\alpha+\beta+3}(4c_d)^{2\alpha+\beta}d^{(2\alpha+\beta+3)p}\varepsilon^{-(2\alpha+\beta)}N_{d,\varepsilon}(3N_{d,\varepsilon})^{2N_{d,\varepsilon}}\varepsilon^{4+\gamma}\varepsilon^{-(4+\gamma)} \\
&\leq 4B^{2\alpha+\beta+3}(4c_d)^{4+2\alpha+\beta+\gamma}d^{(2\alpha+\beta+3)p}N_{d,\varepsilon}(3N_{d,\varepsilon})^{2N_{d,\varepsilon}}\left(\frac{\sqrt{e}(1+2LT)}{\sqrt{N_{d,\varepsilon}-1}}\right)^{(N_{d,\varepsilon}-1)(4+\gamma)}\varepsilon^{-(4+2\alpha+\beta+\gamma)} \\
&\leq 4B^{2\alpha+\beta+3}(4c_d)^{5+2\alpha+\beta}d^{(2\alpha+\beta+3)p}\sup_{n\in\mathbb{N}\cap[2,\infty)}\left[n(3n)^{2n}\left(\frac{\sqrt{e}(1+2LT)}{\sqrt{n-1}}\right)^{(n-1)(4+\gamma)}\right]\varepsilon^{-(4+2\alpha+\beta+\gamma)} \\
&= 4B^{2\alpha+\beta+3}(4c_d)^{5+2\alpha+\beta}d^{(2\alpha+\beta+3)p}\tilde{C}(\gamma)\varepsilon^{-(4+2\alpha+\beta+\gamma)}.
\end{aligned} \tag{140}$$

Combining this with (128) and (129) proves (113). The proof of Theorem 4.1 is thus completed. □

Acknowledgements

This project has been partially supported through the German Research Foundation via research grants HU1889/6-1 (MH) and UR313/1-1 (TK) and via RTG 2131 *High-dimensional Phenomena in Probability - Fluctuations and Discontinuity* (TAN).

References

[1] C. Beck, S. Becker, P. Grohs, N. Jaafari, and A. Jentzen. Solving stochastic differential equations and Kolmogorov equations by means of deep learning. *arXiv:1806.00421*, 2018.

[2] C. Beck, W. E, and A. Jentzen. Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations. *arXiv:1709.05963*, 2017.

[3] S. Becker, P. Cheridito, and A. Jentzen. Deep optimal stopping. *arXiv:1804.05394*, 2018.

[4] J. Berner, P. Grohs, and A. Jentzen. Analysis of the generalization error: Empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *arXiv:1809.03062*, 2018.

[5] W. E, J. Han, and A. Jentzen. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in Mathematics and Statistics*, 5(4):349–380, 2017.

[6] W. E, M. Hutzenthaler, A. Jentzen, and T. Kruse. Linear scaling algorithms for solving high-dimensional nonlinear parabolic differential equations. *arXiv:1607.03295*, 2016.

[7] W. E, M. Hutzenthaler, A. Jentzen, and T. Kruse. On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations. *arXiv:1708.03223*, 2017.

[8] W. E and B. Yu. The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics*, 6(1):1–12, 2018.

[9] D. Elbrächter, P. Grohs, A. Jentzen, and C. Schwab. DNN Expression Rate Analysis of High-dimensional PDEs: Application to Option Pricing. *arXiv:1809.07669*, 2018.

[10] M. Fujii, A. Takahashi, and M. Takahashi. Asymptotic Expansion as Prior Knowledge in Deep Learning Method for high dimensional BSDEs. *arXiv:1710.07030*, 2017.

[11] P. Grohs, F. Hornung, A. Jentzen, and P. von Wurstemberger. A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *arXiv:1809.02362*, 2018.

[12] J. Han, A. Jentzen, and W. E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.

[13] P. Henry-Labordere. Deep Primal-Dual Algorithm for BSDEs: Applications of Machine Learning to CVA and IM. *Available at SSRN: https://ssrn.com/abstract=3071506*, 2017.

[14] M. Hutzenthaler, A. Jentzen, T. Kruse, T.A. Nguyen, and P. von Wurstemberger. Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations. *arXiv:1807.01212*, 2018.

[15] M. Hutzenthaler and T. Kruse. Multi-level Picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities. *arXiv:1711.01080*, 2017.

[16] A. Jentzen, D. Salimova, and T. Welti. A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients. *arXiv:1809.07321*, 2018.

[17] Y. Khoo, J. Lu, and L. . Ying. Solving parametric PDE problems with artificial neural networks. *arXiv:1707.03351*, 2017.

[18] S. Mishra. A machine learning framework for data driven acceleration of computations of differential equations. *arXiv:1807.09519*, 2018.

[19] M. A. Nabian and H. Meidani. A Deep Neural Network Surrogate for High-Dimensional Random Partial Differential Equations. *arXiv:1806.02957*, 2018.

[20] P. Petersen and F. Voigtlaender. Optimal approximation of piecewise smooth functions using deep ReLU neural networks. *arXiv:1709.05289*, 2018.

[21] A. Pinkus. Approximation theory of the MLP model in neural networks. *Acta Numerica*, 8:143–195, 1999.

[22] M. Raissi. Forward-Backward Stochastic Neural Networks: Deep Learning of High-dimensional Partial Differential Equations. *arXiv:1804.07010*, 2018.

[23] J. Sirignano and K. Spiliopoulos. DGM: A deep learning algorithm for solving partial differential equations. *arXiv:1708.07469*, 2017.